Development of Group Contributionplus Models for Properties of Organic Chemical Systems

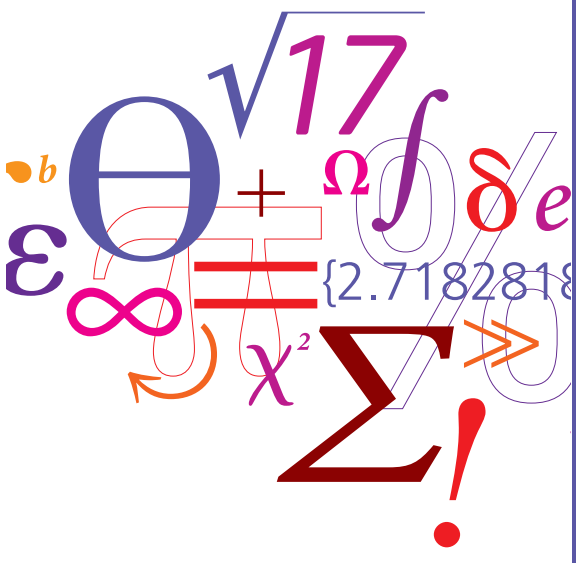
Gonzalez Villalba, Hugo Edson

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

Development of Group Contribution Plus Models for Properties of Organic Chemical Systems



Hugo E. González Villalba
2009

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

Development of Group Contribution^{Plus} Models for Properties of Organic Chemical Systems

Ph. D. Thesis

Hugo E. González Villalba

15 May 2009

Computer Aided Process-Product Engineering Center
Department of Chemical and Biochemical Engineering
Technical University of Denmark

Preface

This thesis is submitted as a partial fulfillment of the requirements for the PhD degree at the Technical University of Denmark ('Danmarks Tekniske Universitet'). The work was done in CAPEC at the Department of Chemical & Biochemical Engineering ('Institut for Kemiteknik') from October 2005 to December 2008 under the supervision of Prof. Rafiqul Gani and Assoc. Prof. Jens Abildskov. I would like to express my gratitude to Prof. Rafiqul Gani for his guidance, academic support and interest in my work. Also, I am grateful to Assoc. Prof. Jens Abilskov for all the fruitful discussions besides the support provided for the computational calculations. Financial support from CONACyT (Consejo Nacional de Ciencia y Tecnología) is acknowledged, without this economic support this work would not be possible.

I would like to thank all the personnel in CAPEC that offered their help and support whenever it was needed. I would also like to thank all my friends in CAPEC with whom I had fruitful discussions and shared so many things (good most of the time). Special thanks are reserved for Brice Le Bert and Tariq Khan which contributed in crucial stages of this work.

There are many things that stay hidden in a published manuscript; Merlin and Ricardo helped me a lot with those things. Of course, my immense gratitude to my family which gave unconditional support in good and bad times: my parents, Hilda and Gabriel, for raising me and making me who I am now; my brothers and sisters for their support; Belen for her love and support; and Teresita, for showing me how far can you go in the adversity.

Finally, i would like to thank my Milica. Besides many revisions and discussions of the thesis, her love, comprehension and support gave me the strength to accomplish this work.

Summary

Properties of chemicals are fundamental for the design and analysis of chemical, pharmaceutical, food, agrochemical and related industries. In order to meet the increased demand with respect to complexity of the chemical molecular structures, wider range of chemicals and accuracy, further development of existing property estimation methods and techniques and/or development of new models are necessary. The objective of this work is to develop new features for the existing property prediction methods.
The extended estimation methods need to be computationally simple and efficient, so they can be used routinely for process-product engineering calculations. Due to the fact that most of the molecular simulation approaches are far from being applied for thermodynamic modeling in conventional process simulators; thermodynamic models need to be fast and reliable to be used, and at the same time scale as the usual process calculations. Therefore, the group contribution approach, which has been finding increased use in both pure component and mixture property calculations, is considered as the basic model to improve. The main challenge here is to increase the applicability, accuracy and versatility of the group contribution (GC) models without requiring additional experimental data. The aim is the development of a hybrid model that combines molecular descriptors theory and group contribution methods for pure component and for mixture properties prediction. Models of this type are commonly known as GC^{Plus} approach.

The main idea of this methodology is the use of connectivity indices (CI) to describe the molecular fragmentation that is characteristic for the UNIFAC group contribution method and that relates properties (molecular and atomic interactions in this case) with molecular structure. The result is the automatic generation of group interaction parameters (GIPs) for the UNIFAC group contribution method.

A brief introduction to this work, together with some useful thermodynamic relations, is shown in Chapter 1. The objectives and the motivation behind the work are also discussed in Chapter 1.

The background related to the theoretical development of the GC^{Plus} models is given in Chapter 2. Important information related to the use of molecular descriptors combined with traditional group contribution methods is highlighted.

The model development of various GC^{Plus} models corresponding to their reference UNIFAC models is described in Chapter 3. The original UNIFAC-CI (VLE and LLE) methods, together with the modified UNIFAC-CI (Dortmund) models are derived and the group definitions are defined in terms of stoichiometry and connectivity indices.

Several examples highlighting the relation between group and connectivity indices parameters are shown, and the step-by-step calculation of the group interaction parameters is also given in this chapter. Finally, the application range is discussed with basic rules and recommendations for the user.

Chapter 4 describes the procedure for the calculation of the needed atom interactions (for the back-calculation of UNIFAC group interaction parameters) via parameter optimization. This parameter estimation is based on measured binary data for around 400 systems covering UNIFAC groups, both functional and molecular. The correlation results are considered good in most cases, however the following limitations have to be mentioned:

- Generally, systems involving alcohols and carboxylic acids are more difficult to describe;
- There is not yet a systematic procedure to monitor the interdependency between atom interaction parameters present simultaneously in different UNIFAC group interactions; in other words, a complete knowledge about the way in which a certain system influences a given correlation, is elusive;
- The parameterization of the Original UNIFAC-CI (VLE and LLE) model and UNIFAC-CI (Dortmund) model needs to be extended to cover more UNIFAC groups and increase the prediction range.

In Chapter 5, the connectivity index-based UNIFAC group interaction parameters are tested for binary data sets not used for parameter optimization with good results inside the application range. Difficulties and inconsistencies are discussed. It should be noted that a broader set of diverse systems is needed for validation. Thus, the validity of derived GIPs should not be considered fully tested as yet. Also in this chapter, case studies regarding the application and use of the generated UNIFAC group interaction parameters are discussed through illustrative examples and case studies.

Finally, in Chapter 6, a discussion with concluding remarks and recommendations for future work is presented.

Resumé på Dansk

De fysisk/kemiske egenskaber for kemikalier er fundamentale for design og analyse af kemiske, farmaceutiske, fødevare- og agrokemiske industrier, såvel tilsvarende industrier. For at imødekomme stigende krav om kompleksitet af molekylær-kemiske strukturer, en bredere vifte af kemikaler samt præcision, er det nødvendigt at videreudvikle nuværende metoder til estimering af fysisk/kemiske egenskaber, og/eller udvikle nye modeller. Formålet med dette arbejde er at udvikle nye funktioner i de nuværende metoder. Samtidig skal de udvidede metoder være beregningsmæssigt simple og effektive, så de rutinemæssigt kan anvendes i procesmæssige sammenhænge.

Da molekylsimuleringer er langt fra anvendelse i termodynamisk modellering i konventionelle processimulatorer, er det nødvendigt at termodynamiske modeller er hurtige og pålidelige, hvis de skal anvendes indenfor samme tidsramme som sædvanlige procesbereginger. Gruppebidragsmetoden, som finder større og større anvendelse ved beregninger indenfor både renkomponent- og blandingsegenskaber, betragtes derfor som basismodellen til videreudvikling. Her er den primære udfordring at øge anvendelsen, præcisionen, samt alsidigheden af gruppebidragsmodellerne (GC – 'Group Contribution') uden behov for yderligere eksperimentelt data. Målet er at udvikle en hybrid model, der kominerer molekylær teori og gruppebidragsmetoder til forudsigelse af renkomponent- og blandingsegenskaber. Modeller af denne type kaldes sædvanligvis GC^{plus}-metoder. Centralt i denne metodologi er anvendelsen af konnektivitetsindexes (CI – 'Connectivity Indices') til at beskrive de molekylære fragmenter, der danner basis for UNIFAC gruppebidragsmetoden, og forbinder egenskaber (molekylære og atomare interaktioner i dette tilfælde) med molekylær struktur. Resultatet er en automatisk generering af gruppeinteraktionsparametre (GIP) til UNIFAC gruppebidragsmetoden.

En kort introduktion til dette arbejde, sammen med nogle nyttige termodynamiske relationer, er vist i kapitel 1. Målsætningerne og motivationen bag dette arbejde er ligeledes diskuteret i kapitel 1.

Baggrunden, der relaterer sig til den teoretiske udvikling af GC^{plus}-modeller, er givet i kapitel 2. Vigtig information relateret til anvendelsen af molekylære deskriptorer kombineret med traditionelle gruppebidragsmetoder er fremhævet.

En gennemgang af den indledende indsats mod udviklingen af GC^{plus}-modeller er præsenteret i kapitel 3. De originale UNIFAC-CI (VLE og LLE) metoder, sammen med de modificerede UNIFAC-CI (Dortmund) modeller er udledt og gruppedefinitioner er givet i form af støkiometri og konnektivitetsindexes. Eksempler, fremhævende relationen mellem gruppe- og konnektivitetsindex-baseret

interaktionsparametre er vist trin for trin. Til slut er rammerne for anvendelse diskuteret i forhold til basale regler, samt anbefalinger til brugeren.

Kapitel 4 beskriver proceduren for beregning af nødvendige atomare interaktioner for tilbageberegning af UNIFAC gruppeinteraktionsparametre via parameteroptimering. Denne parameterestimering er baseret på målte data for omkring 400 binære systemer, der dækker både funktionelle og molekylære UNIFAC grupper.
Generelt vurderes resultaterne fra korrelationen som værende tilfredsstillende. Dog bør følgende begrænsninger noteres:

- I de fleste tilfælde er systemer involverende alkoholler samt karboxylsyrer svære at beskrive.
- Der foreligger endnu ikke en systematisk procedure til overvågning af den indbyrdes afhængighed mellem atom-interaktionsparametre i forskellige UNIFAC gruppeinteraktioner. Med andre ord, er en komplet viden om indflydelsen af et givent system på korrelationen flygtig.
- En højere grad af videreudvikling af parameteriseringen for den originale UNIFAC-CI model, og i særdeleshed den modificerede UNIFAC-CI (Dortmund) model, er påkrævet.

I kapitel 5 testes konnektivitetsindex-baserede UNIFAC gruppeinteraktionsparametre på binære datasæt, ej anvendt i parameteroptimeringen, med gode resultater inden for rammerne af anvendelse. Problemer og uoverensstemmelser er diskuteret. Det bør noteres, at en bredere vifte af forskelligartede systemer er nødvendige for validering. Således bør de fundne GIP ikke betragtes som endnu ej værende fuldt testede. Ligeledes i kapitel 5 er case studies omhandlende anvendelsen og brugen af de genererede UNIFAC gruppeinteraktionsparametre diskuteret, gennem illustrative eksempler.

Til slut, i kapitel 6, er en diskussion med afsluttende bemærkninger og fremtidigt arbejde præsenteret.

LIST OF CONTENTS

Chapter 1

Introduction

The knowledge of properties of chemicals is fundamental for the design and analysis of chemical, pharmaceutical, food, agrochemical and related industries. The most common way (and the most reliable) of obtaining physical properties is through experiments. Also, depending on the property under study, a chemical engineer relies on the results of different experimental techniques in order to apply the gained insights at industrial scale. However, due to the increased complexity in the molecular structures of many new compounds (bigger molecules, functional groups that are new and have not been encountered before), the cost and time needed for certain experiments, other alternatives to experiments need to be considered. This should also be true when certain experiments cannot be performed for safety (or hazard) reasons.

Mathematical models have been developed and continue to be developed in order to estimate, correlate and predict thermodynamic properties of pure compounds and mixtures. The main reason for using property models is that they represent an alternative way of obtaining accurate values of thermodynamic properties in a quick and inexpensive manner compared to experiments. However, as time passes, the work of a model developer is getting more challenging due to industrial requirements of properties of new substances that have been discovered or the prediction of complex properties that have become important in a relevant application. The importance of such models in chemical industry is critical since most of the chemical-related companies use these property models embedded into their computer-aided tools (process simulators and/or property prediction toolboxes) for process simulation and screening purposes. Process simulators such as ASPEN, PROII, and HYSYS rely on property models for the simulation of their separation processes. On the other hand, when modeling and designing equipment, these properties are essential for a thorough approach, or simply to take into account non-idealities of the systems under study.

1.1 PHASE EQUILIBRIA: FUNDAMENTALS

In order to define the fundamental criterion for phase equilibrium, Clausius inequality can be taken as a starting point:

$$\Delta U - T\Delta S - W \leq 0 \qquad \text{for a finite change} \tag{1.1}$$

where U is the internal energy, S the entropy and W is work. W can be expressed as $W=W'-P\Delta V$ (where W' is the shaft work, and V is the volume) in order to give:

$$\Delta U - T\Delta S + P\Delta V - W' \leq 0 \tag{1.2}$$

If the system just exchanges work due to volume changes with the surroundings at constant pressure, then the following can be written:

(1.3)

$$\Delta U - T\Delta S + P\Delta V \leq 0$$

that, combined with the Gibbs free energy definition:

$$G = U + PV - TS \tag{1.4}$$

gives the fundamental criterion for thermodynamic phase equilibrium:

$$\Delta G_{T,P} \leq 0 \tag{1.5}$$

When information is needed related to phase equilibrium in a system where composition changes are important, *U* can be expressed as:

$$\forall \;\; i = 1\,,\; NC \qquad\qquad U = U\left(S, V, n_1, n_2, ..., n_i\right) \tag{1.6}$$

where *NC* denotes the number of components in the system, *i* denotes species type and *n* denotes composition. Rewriting this expression in differential form gives:

$$dU = \left(\frac{\partial U}{\partial S}\right)_{V,n_i} dS + \left(\frac{\partial U}{\partial V}\right)_{S,n_i} dV + \sum \left(\frac{\partial U}{\partial n_i}\right)_{S,V,n_j} dn_i \tag{1.7}$$

According to the chemical potential definition:

$$\mu_i = \left(\frac{\partial U}{\partial n_i}\right)_{S,V,n_j} \tag{1.8}$$

which, combined with the equation below (which is the infinitesimal version of expression given in Equation (1.6)):

$$dU = TdS - PdV \tag{1.9}$$

takes the following form:

$$dU = TdS - PdV + \sum \mu_i dn_i \tag{1.10}$$

When *d(PV)* and *d(TS)* are added and subtracted, respectively, on both sides of the Equation (1.10), it can be rewritten as:

$$dG = VdP - SdT + \sum \mu_i dn_i \tag{1.11}$$

Taking into account a closed system with π phases in equilibrium, on constant pressure and temperature, and allowing mass transfer between phases, Equation (1.11) can be rewritten as:

$$dG_{system} = \sum_{k=1}^{\pi} \left[\sum_{i=1}^{C} \mu_i^{(k)} dn_i^{(k)} \right]_{P,T} = 0 \tag{1.12}$$

where k denotes each of the phases, C denotes the number of components, and i denotes the species type.

If a mass balance for each species type is formulated as follows:

$$dn_i^{(1)} = -\sum_{k=2}^{\pi} dn_i^{(k)} \tag{1.13}$$

and this expression substituted in the Equation (1.12), the obtained expression is:

$$\sum_{k=2}^{\pi}\left[\sum_{i=1}^{C}\left(\mu_i^{(k)} - \mu_i^{(1)}\right) dn_i^{(k)}\right] = 0 \tag{1.14}$$

An analysis of Equation (1.14) concludes that:

$$\mu_i^{(1)} = \mu_i^{(2)} = \mu_i^{(3)} = \ldots \mu_i^{(\pi)} \tag{1.15}$$

This means that in an equilibrium state, the chemical potential of species type i must be the same in each of the co-existing phases.

1.1.1 Fugacity as a criteria for phase equilibrium

Since in practice it is difficult to measure chemical potentials, another measurable quantity, namely fugacity, is used for practical purposes. If the fundamental relation for molar volume given below:

$$\left(\frac{\partial \mu}{\partial P}\right)_T = \underline{V} \tag{1.16}$$

is applied to an ideal gas, the following expression can be derived:

$$d\mu = \underline{V} dP = \frac{RT}{P} dP = RT d \ln P \tag{1.17}$$

This expression can be integrated to give:

$$\mu = \mu^{\circ} + RT \ln P \tag{1.18}$$

where μ° is an integration constant that is a function of temperature only.

For a substance in solution, and considering the fugacity, f as a thermodynamic equivalent to pressure, one can write:

$$\overline{\mu}_i = \mu_i{}^{\circ} + RT \ln \overline{f}_i \tag{1.19}$$

or

$$\left[d\overline{\mu}_i = RTd \ln \overline{f}_i \right]_T \tag{1.20}$$

Considering a system of two phases, α and β, integrating the Equation (1.20) and applying the phase equilibrium criterion given in Equation (1.15), gives the following expression:

$$\overline{\mu}_i^{\alpha} - \overline{\mu}_i^{\beta} = 0 = RT \ln \frac{\overline{f}_i^{\alpha}}{\overline{f}_i^{\beta}} \tag{1.21}$$

which demonstrates that the isofugacity criterion can be used to describe phase equilibrium, as below:

$$\overline{f}_i^{\alpha} = \overline{f}_i^{\beta} \tag{1.22}$$

For a vapor-liquid treatment, the vapor phase fugacities can be written in terms of the fugacity coefficients φ_i:

$$f_i^V = y_i \varphi_i^V (T, P, \underline{y}) P \tag{1.23}$$

while the liquid phase fugacities can be represented either in terms of fugacity coeficcients, φ_i or in terms of activity coefficients, γ_i, as given below in Equations (1.24) and (1.25), respectively:

$$f_i^L = x_i \varphi_i^L (T, P, \underline{x}) P \tag{1.24}$$

$$f_i^L = x_i \gamma_i f_i^{\circ} \tag{1.25}$$

where f_i° is the fugacity of species type i in the standard state to which γ_i refers. A vapor-liquid equilibrium approach using both fugacity coefficients is referred to as a φ-φ approach, but if the liquid phase is described using activity coefficients, a φ-γ approach results. The term f_i° in Equation (1.25) is calculated as follows:

$$f_i^{\circ} = P_i^{\circ} \varphi_i^{\circ} POY_i \tag{1.26}$$

where P_i° is the pure component saturation pressure at the system temperature, and φ_i° is the pure component fugacity at saturation. POY_i is the Poynting correction factor.

Whenever the φ-φ approach is chosen, equations of state (EoS) are used for the calculation of phase equilibrium. The use of EoS presents a lot of advantages when calculating phase equilibrium. For example, the full description in the critical region is possible to calculate, and the number of parameters used in the models is reasonable. Also, dealing with non-condensable gases represents no problem (which is not the same for some of the φ-γ models). On the other hand, the use of EoS is not totally successful for the liquid phase when systems with polar molecules are to be handled (here, the φ-γ approach is preferred). In addition, many of these models are capable of important

predictions based on an extrapolation of the structure of the molecules. The work in this PhD thesis takes a direction towards property models under the φ-γ approach.

1.1.2 φ-γ Approach

The activity coefficient, γ, is usually calculated using the excess properties, such as the excess Gibbs energy. Using the expression defining the excess Gibbs energy definition expression as a starting point:

$$\overline{G}_i^E = RT \ln \gamma_i \tag{1.27}$$

and combining the expression with the fundamental relation for the excess properties, given below:

$$d\left(\frac{nG^E}{RT}\right) = \frac{nV^E}{RT}dP - \frac{nH^E}{RT^2}dT + \sum_i \frac{\overline{G}_i^E}{RT}dn_i \tag{1.28}$$

the following expression for activity coefficient at restricted conditions of constant pressure and temperature and $n_j \neq n_i$, is obtained:

$$\ln \gamma_i = \left[\frac{\partial\left(nG^E / RT\right)}{\partial n_i}\right]_{P,T,n_j} \tag{1.29}$$

G^E is a function of temperature, pressure and composition, but for liquids at low to moderate pressures (up to 5 MPa), G^E and its derivatives are very weak functions of pressure. Since these expressions are too general for practical purposes, and in order to make data reduction generating good extrapolations of existing data, several schemes for correlations of measured data exist.

1.2 CORRELATION SCHEMES

Abildskov[1] describes classification of correlation schemes as follows:

Type 1, Gamma infinity
These correlations are based on relationships between molecular structure and nonideality of components in solutions, providing reasonable calculations of activity coefficients at infinity dilution in pure solvent phases. The most popular models are those of Pierroti et al.[2] and Thomas and Eckert[3] (MOSCED and all of its derivations).

Type 2, Composition Effects
These methods are based on correlations allowing estimation of composition effects on the values of infinity dilution quantities, and making these quantities to be used directly for multicomponent systems. Examples of these models are the quasichemical theory of Guggenheim[4] and all of the local composition methods, such as NRTL (Non Random Two Liquids) and UNIQUAC (Universal Quasichemical).

Type 3, Analytical Solutions of Groups Methods
These correlations are based on the principle of groups of atoms representing the molecules. These correlations allow estimations at finite concentrations as well as at infinite dilution. Their range of applicability is wider than for the other methods previously described, and the extrapolation to systems for which there are no experimental data available is possible. These methods are widespread in routine calculation in most of the process simulators available on the market. The most popular of these methods are the Analytical Solution of Groups (ASOG) (Wilson and Deal[5]) and UNIFAC (Fredenslund et al.[6]).

1.3 MOTIVATION AND OBJECTIVES

The work in this PhD thesis deals entirely with activity coefficient models, specifically UNIFAC models, and it is limited to organic systems, which are involve in many of the unit operations and chemical and biochemical processes encountered in industry (such as, oil refinement processes, azeotropic distillation, crystallization, extraction, and pervaporation). Electrolytes, polymers and non-condensable gases are not considered in this work. The UNIFAC models are well established in the chemical engineering literature and used in most of the commercial simulators and chemical engineering software. Since one of the main drawbacks of the UNIFAC method is the need for parameters (approximately 50 % of the parameters are missing in the parameter tables) the purpose of this work is to develop a methodology to generate the missing UNIFAC group interaction parameters without the need for new experimental data and using solely structural information of the groups. The aim is to extend the range of any UNIFAC model by generating the missing parameters. This is done by expressing the UNIFAC parameters a_{mn} as a function of molecular descriptors with the stoichiometry of the atoms playing a role in the calculation. In addition, a minimal amount of adjustable parameters are regressed.

The conventional procedure when filling the UNIFAC parameter tables is performing experiments using compounds involving the missing parameters and then matching the experimental data to the parameters. Even though this is the most reliable way of generating and updating parameters, the procedure is tedious and resource (time and money) consuming. Revisions of UNIFAC parameters have been done every 3 or more years. That is why if a faster, simpler method with reasonable accuracy could be found, the impact of such a method would be of paramount importance for the UNIFAC model users.

This work follows the GC^{Plus} concept proposed by Gani et al.[7] with the idea of taking advantage of already established GC models by modifying and enriching them with complementary information (of molecular and atomic nature) in order to improve the accuracy of their predictions and extend their application range. The GC^{Plus} approach has already been implemented for conventional pure compound property models[7] (for example, extending the application range for GC models for viscosity) and for the prediction of new properties of polymers[8].

Another objective of this work is to develop new features of the current UNIFAC methods so that they can handle increased complexity of molecular structures by enabling the generation of missing parameters and/or allowing the possibility of increased structural representation of new molecules. The proposed models should have

more versatile property model parameter tables without the need for too much additional experimental data (as most available data have already been used by the current models). At the same time, the developed estimation methods need to be simple and efficient, so that they can be successfully used routinely for process-product engineering calculations. Due to the reasons mentioned above, the UNIFAC group contribution approach, which has been finding increasing use for mixture property calculations, is considered as the basic model to improve. The result is the development of a computer-aided framework for mixture properties prediction using a GC^{Plus} approach.

1.4 STRUCTURE OF THE PhD THESIS

The thesis is organized as follows:

- In **Chapter 2**, the necessary background for the understanding of the theoretical development is given. A detailed review and the analysis of the group contribution approach are presented. Also, important limitations are highlighted in order to show opportunities to overcome them. Then, the molecular description theory is discussed in order to introduce the connectivity index concept and its relation with the GC^{Plus} concept. Afterwards, a functional analysis relating the inherent properties of the connectivity indices with fundamental concepts such as molecular volume and electronegativity, is presented. Finally, it is explained in detail how the GC methods are combined with additional molecular information to generate hybrid models.

- A conceptual analysis of the GC^{Plus} concept is given in **Chapter 3** together with the formulation of the UNIFAC-CI models for VLE, LLE and Modified UNIFAC-Dortmund. The derived equations are presented, and the relation between CIs and UNIFAC groups is discussed, as well as the proposed application range. Several examples are formulated in order to clarify the role that the connectivity indices play and, in addition, a statistical analysis for the application range is performed for each of the UNIFAC models.

- In **Chapter 4**, the procedure for the estimation of the atom interaction parameters (AIPs) is described in detail: the minimization technique and objective functions used are defined, as well as descriptions of specific numerical difficulties are discussed, and the way they were sorted out is explained. Correlation analysis, comparisons and regressed parameters that have been developed for the UNIFAC-CI models are given. Several figures with comparisons of different statistical merits (types of deviations or errors) are presented.

- An analysis of the potential predictive power using the CI-methodology is performed in **Chapter 5**. The analysis is done in qualitative and quantitative terms. Special cases and difficulties are highlighted. In addition, a brief discussion and conclusions with recommendations to the user are given for each of the UNIFAC-CI model.

- **Chapter 6** explores the applications of the GC^{Plus} approach through examples and case studies of industrial interest.

- **Chapter 7** gives the conclusions of this work and suggestions for future directions.

Chapter 2

Conceptual Background

2.1 INTRODUCTION

In this chapter, the fundamental background needed to introduce the GC^{Plus} concept for the prediction of thermodynamic properties is given. First, a review of the state of the art for currently available property prediction methods is discussed, followed by a detailed description of the development of GC methods for mixtures and the formulation of the UNIFAC model. The basic concepts of additional molecular information (Connectivity Indices) are then introduced. Connectivity Indices will complement the UNIFAC models for the generation of missing group interaction parameters. Finally, a review of the preliminary efforts when combining GC methods and connectivity indices is presented, followed by the explanation of the link between the UNIFAC model and the connectivity indices.

2.2 STATE OF THE ART IN PROPERTY PREDICTION METHODS FOR ORGANIC SYSTEMS

There are several methodologies and models available for the estimation of useful thermodynamic properties in chemical engineering. Property prediction methods can be categorized, into methods for prediction of pure compounds´ properties and methods for prediction of properties of mixtures. For pure compounds, common properties such as boiling point, melting point, critical properties, glass transition, viscosity, density, thermal conductivity, difussivity, vapour pressure, solubility parameters and many others have been modeled for many years[9]. For mixtures, the most important properties in chemical engineering are phase equilibria properties: Vapor-Liquid Equilibria (VLE), Solid-Liquid Equilibria (SLE), and Liquid-Liquid Equilibria (LLE). Many of the basic unit operations in chemical plants all over the world depend on these properties. For example, Separation processes involving distillation account for more than 50% of the overall separation expenses[10].

Predictive methods are those that allow calculations of properties without previous information or knowledge of the system that is under study. This system can be a mixture at certain temperature and/or pressure or simply a system including one or more new compounds. Focusing on mixtures, this chapter will consider three main types of predictive models: 1) group contribution methods, 2) quantum chemical methods, and 3) QSAR (quantitative structure-activity relationship)/QSPR (quantitative structure-property relationship) methods. In Figure 2.1, a schematic classification of these property prediction models can be seen.

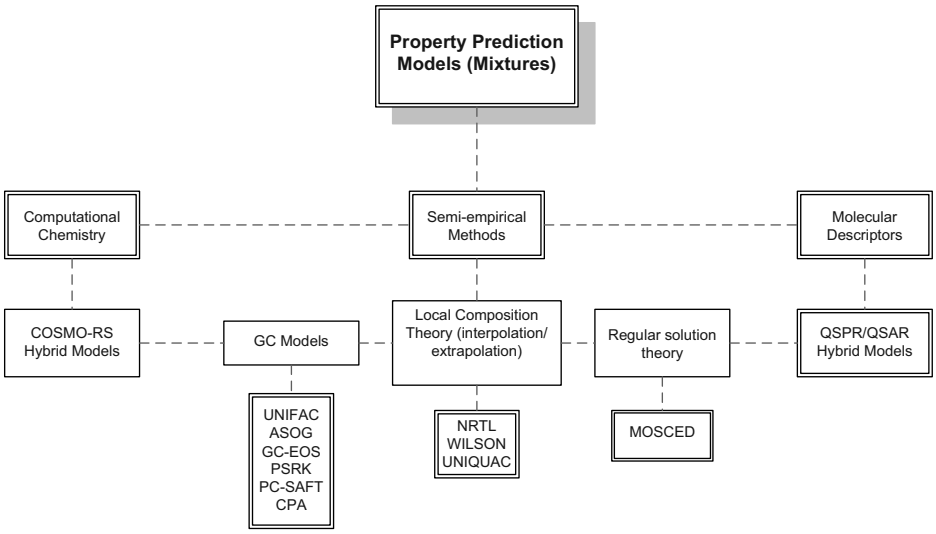


Figure 2.1 Clasification of Predictive Models for Mixture Properties

2.2.1 Group Contribution Methods

The basic assumption behind the group contribution methods is the additivity principle[11] which follows the hypothesis that each of several structural features for a molecular entity makes a separate and additive contribution to a property of the substance concerned. In the case of group contributions, the summation is over the groups representing the chemical under study, as highlighted by Equation (2.1):

$$P_n = \sum_{i=1}^{i=NG} c_i M_i \tag{2.1}$$

where P_n is a specific property '*n*', c_i represents the number of occurrences of the group M_i, and *NG* denotes the total number of groups representing the molecular structure of a chemical.

The formulation of the groups depends on experience and judgment of the model developer. Then, using a reduced number of parameters (the contribution of each group or the interaction between groups), it is potentially possible to describe thousands of compounds (there are 33 million substances registered so far in the Chemical Abstracts Service, CAS[12]), in order to model different types of properties and predict them in an easy and fast way with reasonable accuracy. A detailed analysis of the group contribution models is given in the upcoming sections.

For pure compounds, relatively simple and computationally inexpensive group contribution models for prediction of basic thermodynamic properties have been developed by several authors. Lydersen[13] developed models for critical properties (T_C, P_C, V_C) of organic compounds; and Ambrose[14] also based his work on the development of correlative and predictive models for critical properties. Later on, Klincewicz and Reid[15] developed an improved version of a set of models for critical properties taking

into consideration strengths and weaknesses of previous methods; Joback and Reid[16] presented the first framework for the prediction of nine pure compound properties, including critical properties, normal boiling point and melting point among others; the accuracy and application range were considered very good at the time the method was introduced. Recently, more sophisticated models have arisen in order to improve accuracy and applicability of predictions moving towards a multilevel approach: Constantinou and Gani[17] and Marrero and Gani[18] developed higher order methods where second and third order groups are taken into consideration for larger and functional groups which make a difference in the chemical behavior of the compounds they are representing. This approach also makes possible the distinction between some structural isomers. Good results have been reported using this multilevel approach, both in accuracy and range of applicability (extension to polycyclic compounds).

For mixtures, theoretical efforts have been made for modeling of the non-ideality of the liquid phase using activity coefficients. The group contribution based methods that are most widely used are the UNIFAC[6] group contribution methods. Several modifications of those methods have been proposed over the last decades. Further revisions and extensions of the original UNIFAC, as well as the modifications to UNIFAC are, for example: the two versions of modified UNIFAC developed simultaneously by Larsen et al.[19] and Weidlich et al.[20] for taking into consideration temperature dependence (among other features) - these methods have remained as very successful versions of UNIFAC over time. Another models based on UNIFAC are: DISQUAC[21] (Dispersive Quasichemical model) where each contact between molecules is characterized by a dispersive interchange energy; SUPERFAC[22], a UNIFAC modification that includes temperature dependence and good predictive capabilities for mixtures including large variations in size and shape; and KT-UNIFAC[23] which includes improved correlation and prediction capabilities, distinction between structural isomers and ability to overcome proximity effects. The concept of groups is applied here to describe energetic interactions between the groups describing the molecules of a mixture. These methods require only the structural information of the groups to potentially predict properties of any mixture that can be described using the available groups.

Other advances in prediction of mixture properties were based on development of GC based EoS in order to create a predictive thermodynamic theory, either for the use in a φ-φ approach or in a γ-φ, by combining an EoS with activity coefficient models. Some of the most popular γ-φ approaches are the MVH2-UNIFAC model[24, 25] which is an EoS that uses the existing UNIFAC group interaction parameters allowing VLE predictions at low and high pressure with good accuracy; the predictive Soave-Redlich-Kwong (PRSK)[26] approach combines UNIFAC with the Soave-Redlich-Kwong EoS for predictions over a much larger temperature and pressure range. In the fields of φ-φ approaches, the GC-EoS of Skjold-Jørgensen[27] is based on the generalized van der Waals partition function and uses local-composition mixing rules. It requires only information regarding pure component properties. Also, there are several GC-SAFT[28-32] (Group Contribution-Statistical Associating Fluid Theory) EoS models that are using the GC approach to provide pure compound parameters for the prediction of thermodynamic properties of mixtures.

One of the main drawbacks of the group contribution methods is the dependency of their parameters dependency due to the lack of reliable experimental data. Other limitations of some GC methods are the omission of certain tri-dimensional aspects of

the molecules (for example, stereoisomerism), as well as the difficulty of description of proximity effects. In some cases, the GC methods have a very limited region of applicability.

2.2.2 Quantum Chemical Methods

A more rigorous way of predicting thermodynamic properties is by using quantum chemical methods, such as *Ab initio* method that is based on calculations using first principles which are suppose to require less empirical information; Density Functional Theory (DFT) that determines the properties of electronic systems by using functions of the electron density; Semi-empirical Methods that combine the *Ab initio* calculations with models including parameters matching experimental data; and Molecular Mechanics that use force fields based on Newtonian mechanics to calculate the potential energy of any system. One of the main drawbacks of these methods is the high computational cost for the calculations involved, which makes the calculations impractical for routine chemical engineering calculations and their integration in process simulators. However, some advances include the work of Wolbach and Sandler[33] where they used quantum mechanics methods based on the Density Fuctional Theory and Hartree-Fock for predicting adjustable parameters within the SAFT EoS; good accuracy has been claimed. Recently, Klamt and coworkers[34, 35] have formulated the conductor-like screening model for real solvents (COSMO-RS) which consists of a semi-empirical approach to solve the quantum mechanics calculations (for the surface interaction energies) combined with statistical thermodynamics calculations, in order to predict the needed macroscopic thermodynamic properties. Phase equilibrium calculations are available using COSMO-RS for many mixtures and systems. An issue that should be addressed, however, is the assessment of the model capabilities in predictive mode (since potentially any imaginable mixture can be covered).

2.2.3 The Quantitative Structure-Activity Relationship (QSAR)/Quantitative Structure Property-Relationship (QSPR)

The QSAR/QSPR techniques have been developed with the aim of correlating chemical structures with a defined property or activity. In the case of QSPR, it is assumed that similar molecules have similar properties. The principal applications for these methods are in the area of pure component property correlations, interpolations and extrapolations of vapor pressures, boiling points, melting points, etc. Some work has been done for using QSAR/QSPR techniques for mixture property prediction, work related to the predictions of partition coefficients, absorption, solubilities, and Henry´s law constants among others. The basic steps for the development of QSAR/QSPR models can be summarized as follows:

1) Structure generation;
2) Structure optimization;
3) Descriptor generation;
4) Descriptor reduction;
5) QSAR/QSPR model development; and
6) QSAR/QSPR model validation.

Ravindranath et al.[36] have proposed a hybrid approach integrating excess Gibbs energy (G^E) and QSPR models. This approach is capable of rendering *a priori* predictions for vapor-liquid equilibrium (at different conditions of temperature, pressure and

composition) using NRTL, UNIQUAC and Margules activity coefficient models. A data system of 333 phase equilibrium data sets was used to develop a model parameter database for NRTL, UNIQUAC and Margules methods, and then, using the optimized binary parameters, the QSPR models were developed. The choice of structural descriptors was made by heuristic analysis that was included within the software CODESSA[37]. Nevertheless, this work seems to be of explorative character and further improvement (validation and extension of the application range) is needed. Bunz et al.[38] and Kasturirangan et al.[39] have also conducted studies on developing equations of estate using QSPR modeling.

2.3 GROUP CONTRIBUTION APPROACH FUNDAMENTALS

As pointed out in Section 2.2, the group contribution concept relies on the application of the additivity principle[2] where, as an example, if one considers a carbon atom attached to two hydrogen atoms in a group 'CH_2', it is expected that for a property 'P' of this group will have the same contribution no matter which compound it represents, for example, n-butane molecule or ethylbenzene molecule (see Figure 2.2).

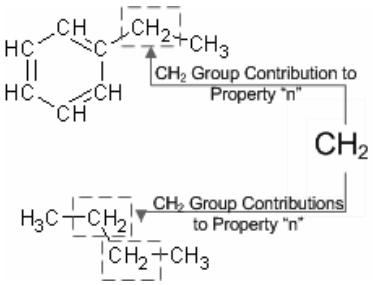


Figure 2.2 Schematical Representation of the Group Contribution Approach

In the context of thermodynamic properties, many attempts for modeling pure component properties have been proposed over the last decades. Most common models include a 'first order approach' of the form:

$$f(x) = \sum_i C_i N_i \tag{2.2}$$

where *f(x)* represents a property function, C_i the number of occurrences, and N_i the group contribution of type *i*.

Many authors have contributed to modeling of thermodynamic properties for pure compounds using the group contribution approach over the years: Lydersen[13] and Ambrose[14] proposed group contribution methods to model critical properties; both methods use group contributions that were obtained by analyzing incremental changes in physical properties within homologous series. Klincewicz[15] developed a method for critical properties (utilizing Ambrose´s data) using a least-squares regression analysis. Benson et al.[40] formulated an approach (intended for the application of thermochemistry in chemistry kinetics) based on the extrapolation of the additivity rules applied for solute species at infinite dilution to model properties such as the enthalpy of formation (ΔH_f), heat capacity (C_p), and entropy (S). Franklin[41] developed a method to estimate

ΔH_f for branched paraffins calculating the group contribution values from the heat content and free energy content functions of molecules containing the groups under study. Verma & Doraiswamy[42] developed a GC method also to estimate ΔH_f as a function of temperature, and for a wider range of organic compounds (hydrocarbons, ethers, ketones, nitrogen-containing, sulfur-containing and halogenated compounds). In this method, the temperature dependence of any group was expressed as a linear relationship, and two equations are required to cover the temperature range of 300° to 1500°C. Van Krevelen & Chermin[43, 44] developed a GC method for the prediction of Gibbs free energy (ΔG_f) as a linear function of temperature directly, however, corrections must be made for molecular symmetry and optical isomers. Thomas[45] developed a method that includes the liquid density and the critical temperature as parameters, and it uses only one coefficient which is determined by a simple scheme of group contributions. The method of Orrick & Erbar[46] estimates the kinematic viscosity as it includes the liquid density at 20°C and the molecular weight determined by group contributions as two additional parameters in their model. For liquid viscosity, the method of Van Velzen et al.[47] uses two parameters that are determined through an extensive group contributions scheme which first requires the calculation of an effective chain length for the given compound. Morris[48] developed a GC method for liquid viscosity which includes only the critical temperature as additional variable and one coefficient that is determined by group contributions. One of the most well known GC methods is the Joback´s[16] framework of GC methods for the prediction of all the properties mentioned above and some more.

'Higher order' methods have been developed in order to increase the range of applicability (properties of large, complex and poly-functional substances). Constantinou & Gani[17] developed a method for several thermodynamic properties using two levels of estimation: level one using the contributions of small first order groups, and level two predicts (corrects) property values by adding the contributions from second and/or higher order groups that take the first order groups as building blocks. The role of the second order groups is to consider, to some extent, the proximity effects and to distinguish among isomers. Later, Marrero & Gani[18] proposed a three-level method for the estimation of nine thermodynamic properties. These methods improve the predictions in accuracy, range of applicability and reliability. The basic level has a large set of simple groups that allows describing a wide variety of organic compounds. However, these groups capture only partially the proximity effects and are unable to distinguish among many isomers. For this reason, the first level of estimation is intended to deal with simple and mono-functional compounds. The second level involves groups that permit a better description of proximity effects and differentiation among isomers. The second level of estimation is consequently intended to deal with poly-functional, polar or non-polar compounds of medium size with number of C-atoms between 3 and 6, and aromatic or cyclo-aliphatic compounds with only one ring and several substituents. The third level has groups that provide more structural information about molecular fragments of compounds whose description is insufficient through the first and second level groups. The third level of estimation allows estimation of complex heterocyclic and large compounds with number of C-atoms between 7 and 60, and poly-functional acyclic compounds.

A typical model for this 'higher order' approach is the following:

$$f(x) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \qquad (2.3)$$

where *f(x)* represents a property *x*, and *N* is the number of occurrences of the first order group *C* or type *i*, while *M* is the number of occurrences of the group contribution *D* of type *j*, and *O* is the number of occurrences of the group contribution E of type *k*.

Probably the first documented attempt to use the group contribution concept for mixtures was done by Irving Langmuir[49] in his work referred to the distribution and orientation of molecules. Langmuir noted that the addition of each successive CH_2 group to a hydrocarbon chain has about the same effect on volume, boiling point and solubility and suggested that it is reasonable to assume that the field of force about any particular group or radical in a large organic molecule is characteristic of that group and, as a first approximation, is independent of the nature of the rest of the molecule. He referred to this as the principle of independent surface action. Of course, if the contributions of the constituent groups (within a molecule) to the free energy of a transfer process are independent, they are also additive. This kind of approach did not receive enough attention until Derr and coworkers (Redlich, Derr and Pierrotti[50]; and Derr and Papadopoulos[51]) developed a group contribution model for prediction of the heat of mixing of non-electrolytes for binary solutions of hydrocarbons. This is a simple theory for the heat of mixing and it is based on the short-range character of intermolecular forces. When developing their model, Derr and coworkers assumed that the energy of interaction of two molecules is the sum of the terms contributed by the contacts between parts of the two molecules. They also pointed out the need that the contacts between different classes (e.g. CH_2 and CO) should have more weight than contacts of groups of the same class (e.g. CH_2 and CH_2). It is assumed, however, that the contributions of each contact depend only on the two groups and not on the other parts and/or the concentration of the two molecules involved or on any other molecules present. Based on this reasoning, they accounted only for interactions of pairs even though it might be considered a crude approximation. Good predictions of the relative partial molal heat contents were found for cases comprising dilute binary systems exhibiting wide ranges of deviation from ideality.

The model and the ideas formulated by Derr and coworkers can be considered as a basis for the development of successive GC methods for mixtures. These ideas together with the good results encouraged Derr and Deal[5] to develop their Analytical Solution of Groups (ASOG) method for correlating and predicting activity coefficients. Parameters associated with pairs of structural groups generated from minimal experimental results are used to calculate activity coefficients with changes in both structure and composition. Their work was tested for many mixtures, which may include combinations from CH_2, OH, O, CO, CN, Cl, and F groups. The model is considered as a sum of two terms:

$$\log \gamma_i = \log \gamma_i^{FH} + \log \gamma_i^{G} \qquad (2.4)$$

where $\log \gamma_i^{FH}$ is the size (combinatorial) contribution and $\log \gamma_i^{G}$ is associated with the interaction of the structural groups (a residual contribution) of the molecule with those of its environment. The size term (γ_i^{FH}) is calculated following the Flory-Huggins

theory. This size term includes a linear functionality that expresses the ratio of solute groups to the total number of groups in the average liquid molecule.

Ratcliff et al.[52] also formulated a similar approach for predicting the excess free energies of liquid mixtures. Later, Kojima and Tochigi[53] increased the range of application of the ASOG approach for a wider range of compounds by adding more group parameters.

2.3.1 The UNIFAC Group Contribution Method

The original UNIFAC (UNIQUAC Functional-group Activity Coefficients)[6] group contribution method combines the ASOG approach with a model for activity coefficients based on an extension of the quasi chemical theory of liquid mixtures (UNIQUAC[54]). In the UNIFAC method, the combinatorial contribution is estimated by using the UNIQUAC equation that contains differences in size and shape of the molecules in the mixture, in addition, functional group sizes and interaction surface areas are introduced from molecular structure data for pure compounds.

In this subsection, only a brief description of the original UNIFAC equations is given (a detailed derivation can be found elsewhere[6, 55]). Those equations are as follow:

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \tag{2.5}$$

where superscript *C* indicates the combinatorial contribution and superscript *R* indicates the residual contribution, γ_i is the activity coefficient of compound *i* in the liquid mixture. The combinatorial term is given by:

$$\ln \gamma_i = \ln \frac{\Phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\theta_i}{\Phi_i} + l_i - \frac{\Phi_i}{x_i} \sum_j x_j l_j \tag{2.6}$$

and:

$$l_i = \frac{z}{2}(r_i - q_i) - (r_i - 1) \tag{2.7}$$

$$z = 10$$

$$\theta_i = \frac{q_i x_i}{\sum_j q_j x_j} \tag{2.8}$$

$$\Phi_i = \frac{r_i x_i}{\sum_j r_j x_j} \tag{2.9}$$

In the above equations, x_i is the mole fraction of component *i*, and the summations included in Equations (2.8) and (2.9) are over all components, including component *i*; θ_i is the area fraction, and Φ_i is the segment fraction which is similar to the volume fraction. Pure component parameters r_i and q_i are, respectively, measures of molecular van der Waals volumes and molecular surfaces areas and are calculated as the sum of the group volume and area parameters, R_k and Q_k:

$$r_i = \sum_k \nu_k^{(i)} R_k \qquad \text{and} \qquad q_i = \sum_k \nu_k^{(i)} Q_k \tag{2.10}$$

where $\nu_k^{(i)}$ is the number of groups (an integer) of type k in molecule i. Using the theory given by Bondi[56], the group parameters R_k and Q_k are obtained as follows:

$$R_k = \frac{V_{wk}}{15.17} \qquad \text{and} \qquad Q_k = \frac{A_{wk}}{\left(2.5 \cdot 10^9\right)} \tag{2.11}$$

where V_{wk} and A_{wk} are the van der Waals group volume and surface area respectively. The normalization factors 15.17 cm^3/mol and $2.5x10^9$ cm^2/mol are those given by Abrams and Prausnitz[54].

The residual term for the activity coefficient term, $\underline{\ln}\, \gamma_i^R$ is given by:

$$\ln \gamma_i^R = \sum_k \nu_k^{(i)} \left[\ln \Gamma_k - \ln \Gamma_k^{(i)} \right] \tag{2.12}$$

where Γ_k is the group residual activity coefficient, and $\Gamma_k^{(i)}$ is the residual activity coefficient of group k in a reference solution containing only molecules of type i. The group activity coefficients Γ_k and $\Gamma_k^{(i)}$ are given by:

$$\ln \Gamma_k = Q_k \left[1 - \ln\left(\sum_m \Theta_m \Psi_{mk} \right) - \sum_m \Theta_m \Psi_{km} / \sum_n \Theta_n \Psi_{nm} \right] \tag{2.13}$$

In Equation (2.13), Θ_m is the area fraction of group m, and the sums are over all the different groups, m is calculated as follows:

$$\Theta_m = \frac{Q_m X_m}{\sum_n Q_n X_n} \tag{2.14}$$

where X_m is the mole fraction of group m in the mixture. The term ψ_{nm} is the group interaction parameter and is given by:

$$\Psi_{mn} = \exp - \left(\frac{U_{mn} - U_{nm}}{RT} \right) = \exp - \left(\frac{a_{mn}}{T} \right) \tag{2.15}$$

In Equation (2.15), U_{mn} is a measure of the energy of interaction between groups m and n. It should be pointed out that a_{mn} has the units of degrees Kelvin and that $a_{mn} \neq a_{nm}$.

Like other models, UNIFAC has some limitations:

- UNIFAC is unable to distinguish between isomers;
- When the γ-φ approach (Equation (1.25)) is used to predict VLE, the vapor phase is usually assumed to be ideal or φ is estimated using the virial equation of

state truncated after the second term. That is why UNIFAC is limited to pressures below 10-15 atm;

- There is a certain temperature range that limits the method (it depends on the range of temperatures used for regressing the group interaction parameters);
- Non-condensable gases are not included;
- Electrolytes are not included;
- There is a lack of group interaction parameters in the UNIFAC parameter tables due to the non availability of experimental data (used to regress the group interaction parameters).

In this work, the focus is on this last limitation of the UNIFAC method. The basic idea is to use a moderate amount of the currently published experimental data together with the molecular descriptors describing the UNIFAC groups to generate the missing UNIFAC group interaction parameters.

2.4 MOLECULAR DESCRIPTION THEORY

Defined by Todeschini[57] as the '*final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment*', molecular descriptors are highly interdisciplinary and have been embraced by the chemical engineering community due to their simplicity when used in routine engineering calculations. Among the most useful molecular descriptors, the connectivity indices-CI (which were introduced in several stages by Randić[58], Kier et al.[59], and Murray et al.[60, 61]) are well known for their ability in correlating physicochemical properties of substances such as boiling point, partition coefficients, molecular refraction, heat of atomization, solubilities, toxicities and many more[62].

2.4.1 Valence Connectivity Indices

While original CIs were intended to deal with molecules containing only carbon, oxygen and hydrogen atoms, Kier-Hall[63] modified the connectivity indices in such a manner so that they were able to include other atoms as well. This new kind of CIs was defined as 'valence connectivity indices' ($^{n}\chi^{v}$). These CIs include several orders that make it possible to describe larger molecular fragments if necessary.

A basic consideration in this approach is that an atom in its valence state can be described by two cardinal numbers, δ and δ^{v}, where δ is the number of bonded neighbors (excluding hydrogen), and δ^{v} is the number of valence electrons (excluding those bonding hydrogen). Once these numbers are defined for each atom, the zeroeth order CI is given by:

$$^{0}\chi^{v} = \sum_{i} \left(\delta_i^{v}\right)^{-1/2} \tag{2.16}$$

where *i* is the atom vertex in the molecule under study.

The first order valence connectivity index is defined as:

$$^{1}\chi^{v} = \sum_{k}\left(\delta_i^{v}\delta_j^{v}\right)_{k}^{-1/2} = \sum_{k} C_k \tag{2.17}$$

where $C_k = (\delta_i^{v}\,\delta_j^{v})_k^{-1/2}$ and *k* is each bond over the entire molecule.

The same procedure is used to derive the second order valence connectivity index

$$^{2}\chi^{v} = \sum_{m}\left(\delta_i^{v}\delta_j^{v}\delta_k^{v}\right)_{m}^{-1/2} = \sum_{m} D_m \tag{2.18}$$

where $D_m = (\delta_i^{v}\,\delta_j^{v}\,\delta_k^{v})_m^{-1/2}$.

Significance of Valence Connectivity Indices $^{n}\chi^{v}$

Kier and Hall[63] introduced the valence connectivity indices in order to account for the heteroatom content in molecules. They encoded in their $^{n}\chi^{v}$ values, fundamental information about volume and electronic character. The 'electronic identity' of each of the different atoms present in a molecule or a group is expressed by the values δ and δ^{v} that are related through the following expressions:

$$\delta_i^{v} = \delta_i + p_i + n_i = \sigma_i + p_i + n_i + h_i \tag{2.19}$$

$$\delta^{v} = Z^{v} - h_i \tag{2.20}$$

In Equation (2.20), Z^{v} accounts for the number of valence electrons not involved in bonds to hydrogen, and h_i is the number of bonded hydrogen on atom *i*.

Z^{v} can be decomposed as $Z^{v} = \sigma_i + p_i + n_i$, where σ_i, p_i, and n_i are related to, respectively, the σ and the π orbital electrons, and lone-pair electrons of each atom. The numbers σ_i, p_i, and *lone-pair* electrons are related to the volume of the constitutent atoms. That is why a trend has been found between the numerical values of $^{n}\chi^{v}$ and Bondi volumes[56]. It is well known that the molecular index ($^{n}\chi^{v}$) is significantly correlated to experimental or calculated volumes using different methodologies. A feature of the molecular index ($^{n}\chi^{v}$) is that the volume will decrease as the numerical values of $^{n}\chi^{v}$ increases.

Electronegativity is important information that can be obtained from CIs, and from electronegativity equalization, properties such as ionicity, bond dipole, partial atomic charge, and bond strength can be derived[64, 65, 66]. Kier and Hall[63] generated successful correlations involving δ^{v} (the fundamental quantity of the CI) with Mulliken electronegativity values.

Functional Analysis

In order to make an analysis of the equations defining the valence connectivity indices and its relation with volume and electronegativity properties, the first order valence connectivity value is used, for example:

$$^{1}\chi^{v} = \sum \left(\delta_i^v \delta_j^v \right)^{-1/2} \tag{2.21}$$

From here, Kier and Hall developed some statements[63] on the rational derivation of these expressions:

- *'The properties of electronegativity and volume encoded in δ_A^v for an atom (A) are considered to be contributed equally from all valence electrons or their orbitals. Thus, one electron from atom A forming one σ bond will possess the fraction 1/ δ_A^v on the properties encoded in δ_A^v.* It can be pointed out that the same is expected from the contribution of atom B, or C for a second order CI.

- *'If the property under consideration is electronegativity, then orbitals from A and B contributing the fractions $1/\delta_A^v$ and $1/\delta_B^v$, respectively, will become adjusted toward an equal intermediate electronegativity, which may be ascribed to the bond'.* Kier and Hall highlight that the expression $[(1/\delta_A^v)\ (1/\delta_B^v)]^{1/2}$ proposed by Sanderson[64, 65] (as an appropriate algorithm for two bonded atoms), is identical to Equation (2.17). Sanderson[66] validated his hypothesis (geometric mean of the atom electronegativities is well correlated with bond electronegativity) in a fruitful discussion on the calculations of polar bond energies via experiments and from electronegativity and related concepts derived on his work.

- '*The volume contributed by atom A to a bond A-B is the fraction $1/\delta_A^v$ of the total. The fraction $1/\delta_B^v$ is the volume contributed from atom B. the bond has a 'volume' due to orbital overlap that can be approximated by $2[(1/\delta_A^v)\ (1/\delta_B^v)]^{1/2}$. With a constant of 2, $[(1/\delta_A^v)\ (1/\delta_B^v)]^{1/2}$ reflects the relative volume of the bond A-B'.* This expression is found to be the same describing electronegativity as was mentioned above.

2.5 HYBRID MODELS: GROUP CONTRIBUTION AND COMPLEMENTARY INFORMATION

2.5.1 Preliminary Efforts

In their pioneering work, Gani et al.[7] established a methodology for predicting missing group contributions for the Marrero and Gani group contribution method, with the aid of valence connectivity indices. The scope of the work of Gani et al.[7] is to use the methodology for any pure component property estimation in the sense of predicting a group contribution that is missing due to limited experimental data. Some rules have to be set in order to describe those groups in terms of connectivity indices. Considering a missing group contribution in a GC model for a property *Y*, the following equations are used in order to predict the property values by generating the missing group contribution:

$$f(Y_m) = \sum_i \left(a_{m,i} A_{m,i} \right) + b\left({}^{v}\chi^{0} \right)_m + 2c\left({}^{v}\chi^{0} \right)_m \tag{2.22}$$

$$f\left(Y^*\right) = \left(\sum_m n_m f\left(Y_m\right) \right) + d \tag{2.23}$$

Where Y is the pure component property to estimate, m indicates the number of different missing group/fragments, n_m indicates the number of times a missing group/fragment appears in the molecule, $A_{m,i}$ is the number of atoms of type i occurring within a missing group/fragment s^*, ${}^{0}\chi^{v}$ and ${}^{1}\chi^{v}$ are (respectively) the zeroeth and first order connectivity indices, $a_{m,i}$ is the contribution of atom i within a missing group/fragment s^*, b and c are adjustable parameters and d is a constant.

Given that the CIs values can be calculated in a straightforward manner (as it will be shown in Chapter 3), in Equations (2.22) and (2.23), the only unknowns left are the parameters b and c which are obtained using a multi-linear regression approach by matching experimental data for ten pure component properties. Each property will have its own set of parameters b and c. Once the parameters are available, the value Y_m, that corresponds to the contribution of a missing group/fragment s^* to property Y, can be calculated. Once Y_m is available, it can be used to estimate the value of the property Y by using the corresponding group contribution model for Y with the added contribution of Y^* as follows:

$$f(Y) = \underbrace{\left(\sum_i N_i C_i \right)}_{\text{GC method}} + f\left(Y^*\right) + \text{higher-order terms} \tag{2.24}$$

One of the main contributions of this work is that this GC-CI methodology has been formulated in such a way that it can be potentially applied to any pure component group contribution method. Currently under the same GC$^{\text{Plus}}$ approach scheme[8] models are developed for pure component properties such as viscosity, surface tension and properties of polymer repeat units.

In Chapter 3, the GC$^{\text{Plus}}$ concept will refer to the use of information from the molecular descriptor theory in order to improve, complement, and further develop group contribution methods.

2.5.2 Relating Group Contribution interactions and Connectivity Indices

The key variables involved in the estimation of the group interaction parameters (GIPs) are the UNIFAC groups which represent the molecules for the system under study. In an atomistic scale, these groups are constituted by atoms and the bonds linking them that give a very specific chemical character to each group. In Section 2.4 it was demonstrated that the CIs translate this atomistic representation into a number encoding important features of the molecules. This means that those numbers can be inferred for each of the UNIFAC groups using CIs, and that each of the UNIFAC groups can then be characterized by the CIs.

Besides the fundamental chemical information being encoded in the CIs, these molecular descriptors have been selected also due to their simplicity, bi-dimensional approach, heteroatomic approach, and their usefulness when implementing them in routine engineering calculations. The groups related to a group interaction a_{ij} are described using valence connectivity indices of three orders: zeroeth order ${}^{0}\chi^{v}$, first order ${}^{1}\chi^{v}$, second order ${}^{2}\chi^{v}$ and the stoichiometry involved in these groups. Basically the group interaction parameter (a_{ij}) is made a function of the CIs and the number of atoms forming the group:

$$a_{ij} = f\left({}^{0}\chi^{v}, {}^{1}\chi^{v}, {}^{1}\chi^{v}, n_{atoms}\right) \tag{2.25}$$

Section 3.3.2 illustrates the step-by-step procedure for representing the UNIFAC groups in terms of the CIs and stoichiometric values.

It has been decided not to use higher than second order connectivity indices, because most of the groups in the UNIFAC parameter tables can be adequately described with segments of two bonds (${}^{2}\chi^{v}$) as a maximum and also because it is important to keep the calculations at a manageable level. When making an inspection of the UNIFAC groups, it can be seen that some of the groups (e.g. CNH group) are small enough not to have CIs of second order.

The main hypothesis of this work is that, if the group interactions in UNIFAC could be described or related to meaningful and simple quantities such as the CIs, the relation between the group interactions and CIs would mean that the gaps in the UNIFAC parameters tables can potentially be filled by generating the missing GIPs using CIs, without the need for new experimental data and in an automatic fashion. This is an important feature for process simulation and design, because, for the implemented UNIFAC model, a 'filled' UNIFAC parameter table could be provided.

Chapter 3

GC^{Plus} Models

3.1 INTRODUCTION

In Chapter 2, the basic concepts providing the background information on the GC^{Plus} models were introduced. This chapter is devoted to a detailed description of each of the GC^{Plus} models that have been developed in this work.

3.2 EXISTING GC^{Plus} MODELS

The first documented work related to the GC^{Plus} approach, is that of Gani and coworkers[7]. In Chapter 2, a brief description was given in order to illustrate the preliminary efforts towards the development of hybrid property models based on group contribution methods. Following the same methodology, work has also been done for other pure component properties for polymers[8]: glassy amorphous volume (V_g), rubbery amorphous volume (V_r) at 298.15 K, amorphous volume (V_a) at 298.15 K, crystalline volume (V_c) at 298.15 K, glass transition temperature (T_g), solubility parameter (δ) at 298.15 K, and refractive index (n) at 298.15 K. Lately, a combined GC and CI based model has been developed for the estimation of surface tension and viscosity[67].

On the other hand, exploring the use of other elements to overcome the inability of GC-models to distinguish between steroisomers, steric properties of molecules (based on molecular mechanics theory[68]) have been combined with GC models obtaining qualitatively acceptable results[69].

The present work is devoted to the development of UNIFAC GC^{Plus} models. UNIFAC models have been chosen because they have been proven to be reliable predictive models, computationally efficient, easy to program, and fast in calculations. Furthermore, and in spite of the gaps in the parameter table; they have a wide application range whenever experimental data are not available. They are widespread methods embedded in most commercial simulators (ASPEN, CHEMCAD, PROII, and others). Only in the 2005-2008 period 425 references[70] (journals, patents, and reviews) have been published involving UNIFAC methods either for direct development and/or parameter delivery, or for their use in correlations, modeling, simulations, comparisons, etc. Because of all these reasons, the application of the GC^{Plus} methodology in order to increase the range of applicability of the UNIFAC methods is a very important issue that deserves explorative, feasibility and development efforts.

3.3 UNIFAC-CI: MODEL FORMULATION

In this work, the following UNIFAC models have been selected to apply the GC^{Plus} approach:

1. Original UNIFAC[71];
2. Original UNIFAC-LLE[72];
3. Modified UNIFAC Dortmund[73].

In order to fill in the gaps of the corresponding UNIFAC model parameter tables, the UNIFAC models are combined with valence connectivity indices (CI). The CIs are used to describe the UNIFAC groups and predict their interactions. When estimating the missing UNIFAC groups, the following procedure is applied:
1) a relation is derived between the group interaction parameter a_{mn} and

- the number of C-atoms in each group;
- the number of O-atoms in each group;
- the number of N-atoms in each group;
- the number of Cl-atoms in each group;
- the number of S-atoms in each group;
- the valence connectivity indices of each group;
- atom interaction parameters (AIPs).

2) Existing VLE data are used to regress the atom interaction parameters, while the atom numbers (stoichiometry) and the connectivity indices are obtained directly from the group definition.

(3) Using the regressed atom interaction parameters and the derived relations (see Equations (3.1) and (3.2)), the missing group interaction parameters a_{mn} and a_{nm} are predicted.

The atom interaction parameters are considered at 4 different levels:
Level 1: interaction between 0^{th} and 0^{th} order connectivity indices;
Level 2: interaction between 0^{th} and 1^{st} order connectivity indices;
Level 3: interaction between 1^{st} and 1^{st} order connectivity indices; and
Level 4: interaction between 0^{th} and 2^{nd} order connectivity indices.

Please note that the interaction between 1^{st} and 2^{nd}, 2^{nd} and 2^{nd} orders are neglected because of the relatively small size and structure of the UNIFAC groups (compared to the molecule they represent). For two groups denoted by m and n, where m is less than n, the following two relations apply (counting all possible atoms, and all interactions at all levels):

$$
\begin{aligned}
a_{mn} = {} & b_{C-C}\left(A_{mn}^{C-C}\right)_0 + b_{C-O}\left(A_{mn}^{C-O}\right)_0 + b_{C-N}\left(A_{mn}^{C-N}\right)_0 + b_{C-Cl}\left(A_{mn}^{C-Cl}\right)_0 + b_{C-S}\left(A_{mn}^{C-S}\right)_0 + b_{O-C}\left(A_{mn}^{O-C}\right)_0 + b_{O-O}\left(A_{mn}^{O-O}\right)_0 + b_{O-N}\left(A_{mn}^{O-N}\right)_0 \\
& + b_{O-Cl}\left(A_{mn}^{O-Cl}\right)_0 + b_{O-S}\left(A_{mn}^{O-S}\right)_0 + b_{N-C}\left(A_{mn}^{N-C}\right)_0 + b_{N-O}\left(A_{mn}^{N-O}\right)_0 + b_{N-N}\left(A_{mn}^{N-N}\right)_0 + b_{N-Cl}\left(A_{mn}^{N-Cl}\right)_0 + b_{N-S}\left(A_{mn}^{N-S}\right)_0 + b_{Cl-C}\left(A_{mn}^{Cl-C}\right)_0 \\
& \underbrace{+ b_{Cl-O}\left(A_{mn}^{Cl-O}\right)_0 + b_{Cl-N}\left(A_{mn}^{Cl-N}\right)_0 + b_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_0 + b_{Cl-S}\left(A_{mn}^{Cl-S}\right)_0 + b_{S-C}\left(A_{mn}^{S-C}\right)_0 + b_{S-O}\left(A_{mn}^{S-O}\right)_0 + b_{S-N}\left(A_{mn}^{S-N}\right)_0 + b_{S-Cl}\left(A_{mn}^{S-Cl}\right)_0 + b_{S-S}\left(A_{mn}^{S-S}\right)_0}_{0th-order\ \text{interactions}} \\
& + c_{C-C}\left(A_{mn}^{C-C}\right)_1 + c_{C-O}\left(A_{mn}^{C-O}\right)_1 + c_{C-N}\left(A_{mn}^{C-N}\right)_1 + c_{C-Cl}\left(A_{mn}^{C-Cl}\right)_1 + c_{C-S}\left(A_{mn}^{C-S}\right)_1 + c_{O-C}\left(A_{mn}^{O-C}\right)_1 + c_{O-O}\left(A_{mn}^{O-O}\right)_1 + c_{O-N}\left(A_{mn}^{O-N}\right)_1 \\
& + c_{O-Cl}\left(A_{mn}^{O-Cl}\right)_1 + c_{O-S}\left(A_{mn}^{O-S}\right)_1 + c_{N-C}\left(A_{mn}^{N-C}\right)_1 + c_{N-O}\left(A_{mn}^{N-O}\right)_1 + c_{N-N}\left(A_{mn}^{N-N}\right)_1 + c_{N-Cl}\left(A_{mn}^{N-Cl}\right)_1 + c_{N-S}\left(A_{mn}^{N-S}\right)_1 + c_{Cl-C}\left(A_{mn}^{Cl-C}\right)_1 \\
& \underbrace{+ c_{Cl-O}\left(A_{mn}^{Cl-O}\right)_1 + c_{Cl-N}\left(A_{mn}^{Cl-N}\right)_1 + c_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_1 + c_{Cl-S}\left(A_{mn}^{Cl-S}\right)_1 + c_{S-C}\left(A_{mn}^{S-C}\right)_1 + c_{S-O}\left(A_{mn}^{S-O}\right)_1 + c_{S-N}\left(A_{mn}^{S-N}\right)_1 + c_{S-Cl}\left(A_{mn}^{S-Cl}\right)_1 + c_{S-S}\left(A_{mn}^{S-S}\right)_1}_{1^{st}-order\ \text{interactions}} \\
& + d_{C-C}\left(A_{mn}^{C-C}\right)_2 + d_{C-O}\left(A_{mn}^{C-O}\right)_2 + d_{C-N}\left(A_{mn}^{C-N}\right)_2 + d_{C-Cl}\left(A_{mn}^{C-Cl}\right)_2 + d_{C-S}\left(A_{mn}^{C-S}\right)_2 + d_{O-C}\left(A_{mn}^{O-C}\right)_2 + d_{O-O}\left(A_{mn}^{O-O}\right)_2 + d_{O-N}\left(A_{mn}^{O-N}\right)_2 \\
& + d_{O-Cl}\left(A_{mn}^{O-Cl}\right)_2 + d_{O-S}\left(A_{mn}^{O-S}\right)_2 + d_{N-C}\left(A_{mn}^{N-C}\right)_2 + d_{N-O}\left(A_{mn}^{N-O}\right)_2 + d_{N-N}\left(A_{mn}^{N-N}\right)_2 + d_{N-Cl}\left(A_{mn}^{N-Cl}\right)_2 + d_{N-S}\left(A_{mn}^{N-S}\right)_2 + d_{Cl-C}\left(A_{mn}^{Cl-C}\right)_2 \\
& \underbrace{+ d_{Cl-O}\left(A_{mn}^{Cl-O}\right)_2 + d_{Cl-N}\left(A_{mn}^{Cl-N}\right)_2 + d_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_2 + d_{Cl-S}\left(A_{mn}^{Cl-S}\right)_2 + d_{S-C}\left(A_{mn}^{S-C}\right)_2 + d_{S-O}\left(A_{mn}^{S-O}\right)_2 + d_{S-N}\left(A_{mn}^{S-N}\right)_2 + d_{S-Cl}\left(A_{mn}^{S-Cl}\right)_2 + d_{S-S}\left(A_{mn}^{S-S}\right)_2}_{2^{nd}-order\ \text{interactions}}
\end{aligned}
$$

$$
\begin{aligned}
&+e_{C-C}\left(A_{mn}^{C-C}\right)_3+e_{C-O}\left(A_{mn}^{C-O}\right)_3+e_{C-N}\left(A_{mn}^{C-N}\right)_3+e_{C-Cl}\left(A_{mn}^{C-Cl}\right)_3+e_{C-S}\left(A_{mn}^{C-S}\right)_3+e_{O-C}\left(A_{mn}^{O-C}\right)_3+e_{O-O}\left(A_{mn}^{O-O}\right)_3+e_{O-N}\left(A_{mn}^{O-N}\right)_3\\
&+e_{O-Cl}\left(A_{mn}^{O-Cl}\right)_3+e_{O-S}\left(A_{mn}^{O-S}\right)_3+e_{N-C}\left(A_{mn}^{N-C}\right)_3+e_{N-O}\left(A_{mn}^{N-O}\right)_3+e_{N-N}\left(A_{mn}^{N-N}\right)_3+e_{N-Cl}\left(A_{mn}^{N-Cl}\right)_3+e_{N-S}\left(A_{mn}^{N-S}\right)_3+e_{Cl-C}\left(A_{mn}^{Cl-C}\right)_3\\
&\underbrace{+e_{Cl-O}\left(A_{mn}^{Cl-O}\right)_3+e_{Cl-N}\left(A_{mn}^{Cl-N}\right)_3+e_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_3+e_{Cl-S}\left(A_{mn}^{Cl-S}\right)_3+e_{S-C}\left(A_{mn}^{S-C}\right)_3+e_{S-O}\left(A_{mn}^{S-O}\right)_3+e_{S-N}\left(A_{mn}^{S-N}\right)_3+e_{S-Cl}\left(A_{mn}^{S-Cl}\right)_3+e_{S-S}\left(A_{mn}^{S-S}\right)_3}_{3^{rd}-order\ interactions}
\end{aligned}
\tag{3.1}
$$

While for the mirror parameters a_{nm} (where m is less than n):

$$
\begin{aligned}
a_{nm}=&\overline{b_{C-C}}\left(A_{nm}^{C-C}\right)_0+\overline{b_{C-O}}\left(A_{nm}^{C-O}\right)_0+\overline{b_{C-N}}\left(A_{nm}^{C-N}\right)_0+\overline{b_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_0+\overline{b_{C-S}}\left(A_{nm}^{C-S}\right)_0+\overline{b_{O-C}}\left(A_{nm}^{O-C}\right)_0+\overline{b_{O-O}}\left(A_{nm}^{O-O}\right)_0+\overline{b_{O-N}}\left(A_{nm}^{O-N}\right)_0\\
&+\overline{b_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_0+\overline{b_{O-S}}\left(A_{nm}^{O-S}\right)_0+\overline{b_{N-C}}\left(A_{nm}^{N-C}\right)_0+\overline{b_{N-O}}\left(A_{nm}^{N-O}\right)_0+\overline{b_{N-N}}\left(A_{nm}^{N-N}\right)_0+\overline{b_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_0+\overline{b_{N-S}}\left(A_{nm}^{N-S}\right)_0+\overline{b_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_0\\
&\underbrace{+\overline{b_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_0+\overline{b_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_0+\overline{b_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_0+\overline{b_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_0+\overline{b_{S-C}}\left(A_{nm}^{S-C}\right)_0+\overline{b_{S-O}}\left(A_{nm}^{S-O}\right)_0+\overline{b_{S-N}}\left(A_{nm}^{S-N}\right)_0+\overline{b_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_0+\overline{b_{S-S}}\left(A_{nm}^{S-S}\right)_0}_{0th-order\ interactions}\\
&+\overline{c_{C-C}}\left(A_{nm}^{C-C}\right)_1+\overline{c_{C-O}}\left(A_{nm}^{C-O}\right)_1+\overline{c_{C-N}}\left(A_{nm}^{C-N}\right)_1+\overline{c_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_1+\overline{c_{C-S}}\left(A_{nm}^{C-S}\right)_1+\overline{c_{O-C}}\left(A_{nm}^{O-C}\right)_1+\overline{c_{O-O}}\left(A_{nm}^{O-O}\right)_1+\overline{c_{O-N}}\left(A_{nm}^{O-N}\right)_1\\
&+\overline{c_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_1+\overline{c_{O-S}}\left(A_{nm}^{O-S}\right)_1+\overline{c_{N-C}}\left(A_{nm}^{N-C}\right)_1+\overline{c_{N-O}}\left(A_{nm}^{N-O}\right)_1+\overline{c_{N-N}}\left(A_{nm}^{N-N}\right)_1+\overline{c_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_1+\overline{c_{N-S}}\left(A_{nm}^{N-S}\right)_1+\overline{c_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_1\\
&\underbrace{+\overline{c_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_1+\overline{c_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_1+\overline{c_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_1+\overline{c_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_1+\overline{c_{S-C}}\left(A_{nm}^{S-C}\right)_1+\overline{c_{S-O}}\left(A_{nm}^{S-O}\right)_1+\overline{c_{S-N}}\left(A_{nm}^{S-N}\right)_1+\overline{c_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_1+\overline{c_{S-S}}\left(A_{nm}^{S-S}\right)_1}_{1^{st}-order\ interactions}\\
&+\overline{d_{C-C}}\left(A_{nm}^{C-C}\right)_2+\overline{d_{C-O}}\left(A_{nm}^{C-O}\right)_2+\overline{d_{C-N}}\left(A_{nm}^{C-N}\right)_2+\overline{d_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_2+\overline{d_{C-S}}\left(A_{nm}^{C-S}\right)_2+\overline{d_{O-C}}\left(A_{nm}^{O-C}\right)_2+\overline{d_{O-O}}\left(A_{nm}^{O-O}\right)_2+\overline{d_{O-N}}\left(A_{nm}^{O-N}\right)_2\\
&+\overline{d_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_2+\overline{d_{O-S}}\left(A_{nm}^{O-S}\right)_2+\overline{d_{N-C}}\left(A_{nm}^{N-C}\right)_2+\overline{d_{N-O}}\left(A_{nm}^{N-O}\right)_2+\overline{d_{N-N}}\left(A_{nm}^{N-N}\right)_2+\overline{d_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_2+\overline{d_{N-S}}\left(A_{nm}^{N-S}\right)_2+\overline{d_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_2\\
&\underbrace{+\overline{d_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_2+\overline{d_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_2+\overline{d_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_2+\overline{d_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_2+\overline{d_{S-C}}\left(A_{nm}^{S-C}\right)_2+\overline{d_{S-O}}\left(A_{nm}^{S-O}\right)_2+\overline{d_{S-N}}\left(A_{nm}^{S-N}\right)_2+\overline{d_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_2+\overline{d_{S-S}}\left(A_{nm}^{S-S}\right)_2}_{2^{nd}-order\ interactions}\\
&+\overline{e_{C-C}}\left(A_{nm}^{C-C}\right)_3+\overline{e_{C-O}}\left(A_{nm}^{C-O}\right)_3+\overline{e_{C-N}}\left(A_{nm}^{C-N}\right)_3+\overline{e_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_3+\overline{e_{C-S}}\left(A_{nm}^{C-S}\right)_3+\overline{e_{O-C}}\left(A_{nm}^{O-C}\right)_3+\overline{e_{O-O}}\left(A_{nm}^{O-O}\right)_3+\overline{e_{O-N}}\left(A_{nm}^{O-N}\right)_3\\
&+\overline{e_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_3+\overline{e_{O-S}}\left(A_{nm}^{O-S}\right)_3+\overline{e_{N-C}}\left(A_{nm}^{N-C}\right)_3+\overline{e_{N-O}}\left(A_{nm}^{N-O}\right)_3+\overline{e_{N-N}}\left(A_{nm}^{N-N}\right)_3+\overline{e_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_3+\overline{e_{N-S}}\left(A_{nm}^{N-S}\right)_3+\overline{e_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_3\\
&\underbrace{+\overline{e_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_3+\overline{e_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_3+\overline{e_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_3+\overline{e_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_3+\overline{e_{S-C}}\left(A_{nm}^{S-C}\right)_3+\overline{e_{S-O}}\left(A_{nm}^{S-O}\right)_3+\overline{e_{S-N}}\left(A_{nm}^{S-N}\right)_3+\overline{e_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_3+\overline{e_{S-S}}\left(A_{nm}^{S-S}\right)_3}_{3^{rd}-order\ interactions}
\end{aligned}
\tag{3.2}
$$

In Equations (3.1) and (3.2), $b_{XY}, c_{XY}, d_{XY}, e_{XY}$, $\overline{b_{XY}}, \overline{c_{XY}}, \overline{d_{XY}}, \overline{e_{XY}}$ are the atom interaction parameters (AIPs, values are obtained through regression of VLE data) between atoms of type *X*, and atoms of type *Y* (*X* and *Y* representing different atoms covered by the methodology, on a H-atom free basis); the maximum number of AIPs needed is for the a_{mn} and a_{nm} expressions (together) is 200. $\left(A_{mn}^{XY}\right)_i$ are the coefficients to be used with the corresponding AIPs, that is, at the order *i*, between atoms of type *X*, and atoms of type *Y*, in the interaction between group *m* and group *n*, and which are defined by the following equations for each order:

$$
\left(A_{mn}^{XY}\right)_0=\frac{n_X^{(m)\,\nu}\chi_{(n)}^0-n_Y^{(n)\,\nu}\chi_{(m)}^0}{^{\nu}\chi_{(n)}^0\ {}^{\nu}\chi_{(m)}^0}
\tag{3.3}
$$

$$
\left(A_{mn}^{XY}\right)_1=\frac{n_X^{(m)\,\nu}\chi_{(n)}^1-n_Y^{(n)\,\nu}\chi_{(m)}^0}{^{\nu}\chi_{(n)}^1\ {}^{\nu}\chi_{(m)}^0}
\tag{3.4}
$$

$$\left(A_{mn}^{XY}\right)_2 = \frac{n_X^{(m)\,{}^{v}\chi_{(n)}^{1}} - n_Y^{(n)\,{}^{v}\chi_{(m)}^{1}}}{{}^{v}\chi_{(n)}^{1}\,{}^{v}\chi_{(m)}^{1}} \tag{3.5}$$

$$\left(A_{mn}^{XY}\right)_3 = \frac{n_X^{(m)\,{}^{v}\chi_{(n)}^{2}} - n_Y^{(n)\,{}^{v}\chi_{(m)}^{0}}}{{}^{v}\chi_{(n)}^{2}\,{}^{v}\chi_{(m)}^{0}} \tag{3.6}$$

where $n_X^{(m)}$ is the number of atoms of type *X* in the group *m*; and ${}^{v}\chi_{(m)}^{i}$ is the *i*-th order valence connectivity index for the group *m*.

The derived Equations (3.1) and (3.2) can then be used to back-calculate the group interaction parameters described in Equation (2.15) for the original UNIFAC model. Note that the actual number of atom interaction parameters needed for any missing UNIFAC group interaction parameters is quite small (compared with the amount of missing UNIFAC group interaction parameters) since not all the terms in Equations (3.1) and (3.2) are needed (as will be highlighted in Example 3.1).

3.3.1 Scope and Significance

Several questions can be postulated for this methodology. Note that the UNIFAC-CI models have been further developed with the aim of filling in the gaps in their corresponding model parameter tables. The application of the methodology assumes that it is possible to generate any particular missing GIP in the UNIFAC parameter tables; whenever the necessary atoms have already been parameterized and the connectivity indices plus the stoichiometry of the groups calculated. The GIPs involving a totally new group can also be calculated after generating the necessary information for that group (volume and surface parameters, R_k and Q_k; stoichiometry and connectivity indices values). This guarantees keeping the parameter table updated whenever new groups are added to it as a result of further development of the UNIFAC model.

The basic methodology in order to generate the missing UNIFAC GIPs through the CI method can be summarized as follows:
1) Compile enough and necessary experimental data.
Enough and necessary experimental data means getting the best experimental data, focusing on: a) getting heterogeneous data (different classes of chemical compounds); b) reliable data in terms of literature sources and thermodynamic consistency[74] whenever possible (just for VLE data). Although using a minimum of experimental data is one of the objectives of this work, there is no perfect criterion in order to choose which and how many experimental data sets to use.
2) Compile stoichiometric and connectivity index information.
Once the UNIFAC parameter table has been selected, a compilation of the stoichiometry and connectivity indices values (zeroth, first and second order) is required. These numbers take part in the calculations using the relations derived for a_{mn} (Equations (3.1) and (3.2)).
3) Perform the AIPs optimization.
Values for the set of AIPs are then determined by regression using the experimental database. Different optimization schemes are expected depending on the UNIFAC model under study.

4) Using the derived relations from Equations (3.1) and (3.2), the GIPs can be generated.

3.3.2 Group Definitions and Stoichiometric/Connectivity Indices Tables

Once a UNIFAC model has been selected, certain information should be extracted from the groups, that is, the number of atoms of type *X* (n_x, stoichiometry) and the connectivity index values of zeroth, first and second order. Unless stated otherwise, the CI values are calculated using the delta values (δ^v) given in Kier-Hall´s work[75] (the CIs values for all the UNIFAC groups used in this work are compiled in Appendix A). For example, to extract the necessary information from the UNIFAC group 12, HCOO (the 2D hydrogen representation is given in Figure 3.1), the calculations are as follows:

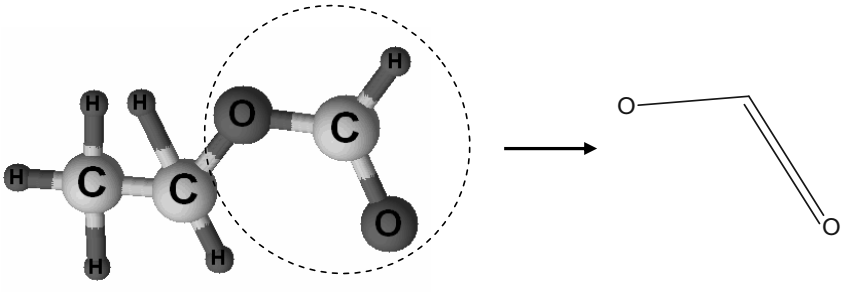


Figure 3.1 UNIFAC Group HCOO and its Representation in a Hydrogen Free Graph

1) The information needed to calculate the delta values (necessary for calculation of the valence connectivity indices) are: the number of hydrogen atoms attached to each vertex-atom (N_H), and the number of valence electrons in the atom (Z^v). Table 3.1 lists the values of δ^v for the atoms involved in the calculation, and Table 3.2 gives the necessary information to calculate the CIs values.

Table 3.1 Atomic Indices Values for the HCOO Group

Atom	1 (oxygen)	2 (carbon)	3 (oxygen)
Z^v	6	4	6
N_H	0	1	0
δ^v	6	3	6

Table 3.2 Calculated Atom, Bond and Triplet Indices for the Group HCOO

Atom	1 (oxygen)	2 (carbon)	3 (oxygen)
δ^v	6	3	6
$1/(\delta^v)^{1/2}$	0.40824829	0.577350269	0.40824829

Bond	a(1-2)	b(2-3)
β^k	18	18
$1/(\beta^k)^{1/2}$	0.23570226	0.23570226

Triplet	a-b (1-2-3)
γ^k	108
$1/(\gamma^k)^{1/2}$	0.096225045

2) Using the calculated values given in Table 3.2, the zeroth order connectivity index $^0\chi^v$, Equation (2.16), is calculated as follows:

$$^{0}\chi^{v} = \sum_{i} \left(\delta_i^{v}\right)^{-1/2}$$

$$^{0}\chi^{v} = \sum_{i} \left(0.40824829 + 0.577350269 + 0.40824829\right)$$

$$^{0}\chi^{v} = 1.393846850117350$$

3) In the same way, using Equations (2.17) and (2.18) together with the the calculated values given in Table 3.2 in Table 3.2, we obtain $^{1}\chi^{v}$= 0.471404521, and $^{2}\chi^{v}$= 0.096225045. The stoichiometric values of this group are: number of carbon atoms n_C=1, and number of oxygen atoms n_O=2.

All the connectivity index values and stoichiometry of the main groups under study are calculated following this procedure. Tables with values of CIs for the main groups of the UNIFAC models can be found in Appendix A.

3.4 UNIFAC-CI FOR VLE

3.4.1 Reference Model (Original UNIFAC-VLE)

The original UNIFAC model (updated by Hansen et al.[71]) has been chosen as the reference model. The parameter matrix of this model contains 50 main groups, and it should be pointed out that more than 50% of the parameters are missing (see Figure 3.2) in the parameter table. Taking into account that the scope of this work is to demonstrate that the GC^{Plus} concept can be successfully applied in a generalized way (form different GC models), Hansen´s parameter table is good enough for the purpose of this work.

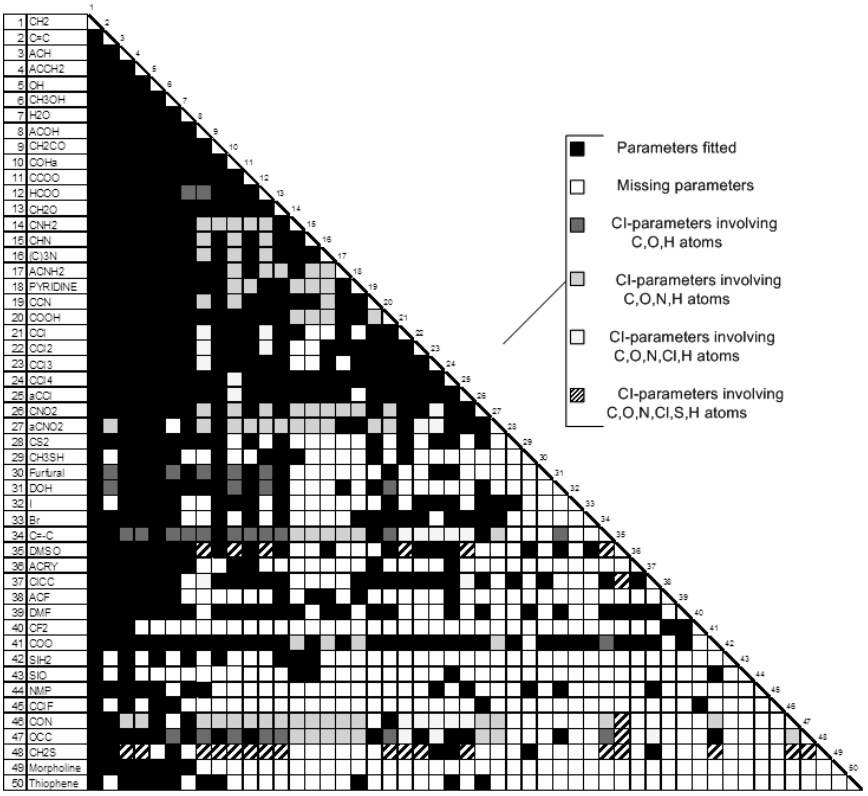


Figure 3.2 UNIFAC Group Interaction Parameter Matrix. Revision 5 [71] *(black squares) + Group Interaction Parameters Generated through Connectivity Indices*

This UNIFAC model is a further extension in the application range and accuracy of the original UNIFAC method published for the first time by Fredenslund et al.[6] and described in detail in Chapter 2. This model is based mainly on VLE data, which means that the model is intended to be used basically for the prediction of VLE systems. However, it has been shown[76] that reasonable predictions of SLE systems can be achieved by using these 'VLE parameters'. In Chapter 5, VLE and SLE predictions using the CI-generated parameters for this original UNIFAC (VLE) model will be discussed.

The model has been formulated to deal with C, O, N, Cl, and S atoms, and the expressions for the group interaction parameters (a_{mn} and a_{nm}) as a function of CI and stoichiometry are given by Equations (3.7) and (3.8). Therefore, the objective of this work is to fill in most of the gaps in the UNIFAC parameter table for the parameter interactions related to groups involving C, O, N, Cl and S atoms. It should be pointed out that parallel parameter tables (for example CH_3OH and H_2O in original UNIFAC) exist for the following molecular groups: CH_3OH, H_2O, PYR, CCl_4, DOH, and DMSO; this means that the GIPs involving these groups should be generated by using their corresponding AIP tables.

The values for the a_{mn} and a_{nm} parameters in original UNIFAC (see Equation (2.15)) can be calculated using the UNIFAC-CI (VLE) model as follows:

$$
\begin{aligned}
a_{mn} = {} & b_{C-C}\left(A_{mn}^{C-C}\right)_0 + b_{C-O}\left(A_{mn}^{C-O}\right)_0 + b_{C-N}\left(A_{mn}^{C-N}\right)_0 + b_{C-Cl}\left(A_{mn}^{C-Cl}\right)_0 + b_{C-S}\left(A_{mn}^{C-S}\right)_0 + b_{O-C}\left(A_{mn}^{O-C}\right)_0 + b_{O-O}\left(A_{mn}^{O-O}\right)_0 + b_{O-N}\left(A_{mn}^{O-N}\right)_0 \\
& + b_{O-Cl}\left(A_{mn}^{O-Cl}\right)_0 + b_{O-S}\left(A_{mn}^{O-S}\right)_0 + b_{N-C}\left(A_{mn}^{N-C}\right)_0 + b_{N-O}\left(A_{mn}^{N-O}\right)_0 + b_{N-N}\left(A_{mn}^{N-N}\right)_0 + b_{N-Cl}\left(A_{mn}^{N-Cl}\right)_0 + b_{N-S}\left(A_{mn}^{N-S}\right)_0 + b_{Cl-C}\left(A_{mn}^{Cl-C}\right)_0 \\
& \underbrace{+b_{Cl-O}\left(A_{mn}^{Cl-O}\right)_0 + b_{Cl-N}\left(A_{mn}^{Cl-N}\right)_0 + b_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_0 + b_{Cl-S}\left(A_{mn}^{Cl-S}\right)_0 + b_{S-C}\left(A_{mn}^{S-C}\right)_0 + b_{S-O}\left(A_{mn}^{S-O}\right)_0 + b_{S-N}\left(A_{mn}^{S-N}\right)_0 + b_{S-Cl}\left(A_{mn}^{S-Cl}\right)_0 + b_{S-S}\left(A_{mn}^{S-S}\right)_0}_{0th-order\ interactions} \\
& + c_{C-C}\left(A_{mn}^{C-C}\right)_1 + c_{C-O}\left(A_{mn}^{C-O}\right)_1 + c_{C-N}\left(A_{mn}^{C-N}\right)_1 + c_{C-Cl}\left(A_{mn}^{C-Cl}\right)_1 + c_{C-S}\left(A_{mn}^{C-S}\right)_1 + c_{O-C}\left(A_{mn}^{O-C}\right)_1 + c_{O-O}\left(A_{mn}^{O-O}\right)_1 + c_{O-N}\left(A_{mn}^{O-N}\right)_1 \\
& + c_{O-Cl}\left(A_{mn}^{O-Cl}\right)_1 + c_{O-S}\left(A_{mn}^{O-S}\right)_1 + c_{N-C}\left(A_{mn}^{N-C}\right)_1 + c_{N-O}\left(A_{mn}^{N-O}\right)_1 + c_{N-N}\left(A_{mn}^{N-N}\right)_1 + c_{N-Cl}\left(A_{mn}^{N-Cl}\right)_1 + c_{N-S}\left(A_{mn}^{N-S}\right)_1 + c_{Cl-C}\left(A_{mn}^{Cl-C}\right)_1 \\
& \underbrace{+c_{Cl-O}\left(A_{mn}^{Cl-O}\right)_1 + c_{Cl-N}\left(A_{mn}^{Cl-N}\right)_1 + c_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_1 + c_{Cl-S}\left(A_{mn}^{Cl-S}\right)_1 + c_{S-C}\left(A_{mn}^{S-C}\right)_1 + c_{S-O}\left(A_{mn}^{S-O}\right)_1 + c_{S-N}\left(A_{mn}^{S-N}\right)_1 + c_{S-Cl}\left(A_{mn}^{S-Cl}\right)_1 + c_{S-S}\left(A_{mn}^{S-S}\right)_1}_{1^{st}-order\ interactions} \\
& + d_{C-C}\left(A_{mn}^{C-C}\right)_2 + d_{C-O}\left(A_{mn}^{C-O}\right)_2 + d_{C-N}\left(A_{mn}^{C-N}\right)_2 + d_{C-Cl}\left(A_{mn}^{C-Cl}\right)_2 + d_{C-S}\left(A_{mn}^{C-S}\right)_2 + d_{O-C}\left(A_{mn}^{O-C}\right)_2 + d_{O-O}\left(A_{mn}^{O-O}\right)_2 + d_{O-N}\left(A_{mn}^{O-N}\right)_2 \\
& + d_{O-Cl}\left(A_{mn}^{O-Cl}\right)_2 + d_{O-S}\left(A_{mn}^{O-S}\right)_2 + d_{N-C}\left(A_{mn}^{N-C}\right)_2 + d_{N-O}\left(A_{mn}^{N-O}\right)_2 + d_{N-N}\left(A_{mn}^{N-N}\right)_2 + d_{N-Cl}\left(A_{mn}^{N-Cl}\right)_2 + d_{N-S}\left(A_{mn}^{N-S}\right)_2 + d_{Cl-C}\left(A_{mn}^{Cl-C}\right)_2 \\
& \underbrace{+d_{Cl-O}\left(A_{mn}^{Cl-O}\right)_2 + d_{Cl-N}\left(A_{mn}^{Cl-N}\right)_2 + d_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_2 + d_{Cl-S}\left(A_{mn}^{Cl-S}\right)_2 + d_{S-C}\left(A_{mn}^{S-C}\right)_2 + d_{S-O}\left(A_{mn}^{S-O}\right)_2 + d_{S-N}\left(A_{mn}^{S-N}\right)_2 + d_{S-Cl}\left(A_{mn}^{S-Cl}\right)_2 + d_{S-S}\left(A_{mn}^{S-S}\right)_2}_{2^{nd}-order\ interactions} \\
& + e_{C-C}\left(A_{mn}^{C-C}\right)_3 + e_{C-O}\left(A_{mn}^{C-O}\right)_3 + e_{C-N}\left(A_{mn}^{C-N}\right)_3 + e_{C-Cl}\left(A_{mn}^{C-Cl}\right)_3 + e_{C-S}\left(A_{mn}^{C-S}\right)_3 + e_{O-C}\left(A_{mn}^{O-C}\right)_3 + e_{O-O}\left(A_{mn}^{O-O}\right)_3 + e_{O-N}\left(A_{mn}^{O-N}\right)_3 \\
& + e_{O-Cl}\left(A_{mn}^{O-Cl}\right)_3 + e_{O-S}\left(A_{mn}^{O-S}\right)_3 + e_{N-C}\left(A_{mn}^{N-C}\right)_3 + e_{N-O}\left(A_{mn}^{N-O}\right)_3 + e_{N-N}\left(A_{mn}^{N-N}\right)_3 + e_{N-Cl}\left(A_{mn}^{N-Cl}\right)_3 + e_{N-S}\left(A_{mn}^{N-S}\right)_3 + e_{Cl-C}\left(A_{mn}^{Cl-C}\right)_3 \\
& \underbrace{+e_{Cl-O}\left(A_{mn}^{Cl-O}\right)_3 + e_{Cl-N}\left(A_{mn}^{Cl-N}\right)_3 + e_{Cl-Cl}\left(A_{mn}^{Cl-Cl}\right)_3 + e_{Cl-S}\left(A_{mn}^{Cl-S}\right)_3 + e_{S-C}\left(A_{mn}^{S-C}\right)_3 + e_{S-O}\left(A_{mn}^{S-O}\right)_3 + e_{S-N}\left(A_{mn}^{S-N}\right)_3 + e_{S-Cl}\left(A_{mn}^{S-Cl}\right)_3 + e_{S-S}\left(A_{mn}^{S-S}\right)_3}_{3^{rd}-order\ interactions}
\end{aligned}
\tag{3.7}
$$

and,

$$
\begin{aligned}
a_{nm} = {} & \overline{b_{C-C}}\left(A_{nm}^{C-C}\right)_0 + \overline{b_{C-O}}\left(A_{nm}^{C-O}\right)_0 + \overline{b_{C-N}}\left(A_{nm}^{C-N}\right)_0 + \overline{b_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_0 + \overline{b_{C-S}}\left(A_{nm}^{C-S}\right)_0 + \overline{b_{O-C}}\left(A_{nm}^{O-C}\right)_0 + \overline{b_{O-O}}\left(A_{nm}^{O-O}\right)_0 + \overline{b_{O-N}}\left(A_{nm}^{O-N}\right)_0 \\
& + \overline{b_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_0 + \overline{b_{O-S}}\left(A_{nm}^{O-S}\right)_0 + \overline{b_{N-C}}\left(A_{nm}^{N-C}\right)_0 + \overline{b_{N-O}}\left(A_{nm}^{N-O}\right)_0 + \overline{b_{N-N}}\left(A_{nm}^{N-N}\right)_0 + \overline{b_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_0 + \overline{b_{N-S}}\left(A_{nm}^{N-S}\right)_0 + \overline{b_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_0 \\
& \underbrace{+\overline{b_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_0 + \overline{b_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_0 + \overline{b_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_0 + \overline{b_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_0 + \overline{b_{S-C}}\left(A_{nm}^{S-C}\right)_0 + \overline{b_{S-O}}\left(A_{nm}^{S-O}\right)_0 + \overline{b_{S-N}}\left(A_{nm}^{S-N}\right)_0 + \overline{b_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_0 + \overline{b_{S-S}}\left(A_{nm}^{S-S}\right)_0}_{0th-order\ interactions} \\
& + \overline{c_{C-C}}\left(A_{nm}^{C-C}\right)_1 + \overline{c_{C-O}}\left(A_{nm}^{C-O}\right)_1 + \overline{c_{C-N}}\left(A_{nm}^{C-N}\right)_1 + \overline{c_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_1 + \overline{c_{C-S}}\left(A_{nm}^{C-S}\right)_1 + \overline{c_{O-C}}\left(A_{nm}^{O-C}\right)_1 + \overline{c_{O-O}}\left(A_{nm}^{O-O}\right)_1 + \overline{c_{O-N}}\left(A_{nm}^{O-N}\right)_1 \\
& + \overline{c_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_1 + \overline{c_{O-S}}\left(A_{nm}^{O-S}\right)_1 + \overline{c_{N-C}}\left(A_{nm}^{N-C}\right)_1 + \overline{c_{N-O}}\left(A_{nm}^{N-O}\right)_1 + \overline{c_{N-N}}\left(A_{nm}^{N-N}\right)_1 + \overline{c_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_1 + \overline{c_{N-S}}\left(A_{nm}^{N-S}\right)_1 + \overline{c_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_1 \\
& \underbrace{+\overline{c_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_1 + \overline{c_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_1 + \overline{c_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_1 + \overline{c_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_1 + \overline{c_{S-C}}\left(A_{nm}^{S-C}\right)_1 + \overline{c_{S-O}}\left(A_{nm}^{S-O}\right)_1 + \overline{c_{S-N}}\left(A_{nm}^{S-N}\right)_1 + \overline{c_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_1 + \overline{c_{S-S}}\left(A_{nm}^{S-S}\right)_1}_{1^{st}-order\ interactions} \\
& + \overline{d_{C-C}}\left(A_{nm}^{C-C}\right)_2 + \overline{d_{C-O}}\left(A_{nm}^{C-O}\right)_2 + \overline{d_{C-N}}\left(A_{nm}^{C-N}\right)_2 + \overline{d_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_2 + \overline{d_{C-S}}\left(A_{nm}^{C-S}\right)_2 + \overline{d_{O-C}}\left(A_{nm}^{O-C}\right)_2 + \overline{d_{O-O}}\left(A_{nm}^{O-O}\right)_2 + \overline{d_{O-N}}\left(A_{nm}^{O-N}\right)_2 \\
& + \overline{d_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_2 + \overline{d_{O-S}}\left(A_{nm}^{O-S}\right)_2 + \overline{d_{N-C}}\left(A_{nm}^{N-C}\right)_2 + \overline{d_{N-O}}\left(A_{nm}^{N-O}\right)_2 + \overline{d_{N-N}}\left(A_{nm}^{N-N}\right)_2 + \overline{d_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_2 + \overline{d_{N-S}}\left(A_{nm}^{N-S}\right)_2 + \overline{d_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_2 \\
& \underbrace{+\overline{d_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_2 + \overline{d_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_2 + \overline{d_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_2 + \overline{d_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_2 + \overline{d_{S-C}}\left(A_{nm}^{S-C}\right)_2 + \overline{d_{S-O}}\left(A_{nm}^{S-O}\right)_2 + \overline{d_{S-N}}\left(A_{nm}^{S-N}\right)_2 + \overline{d_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_2 + \overline{d_{S-S}}\left(A_{nm}^{S-S}\right)_2}_{2^{nd}-order\ interactions} \\
& + \overline{e_{C-C}}\left(A_{nm}^{C-C}\right)_3 + \overline{e_{C-O}}\left(A_{nm}^{C-O}\right)_3 + \overline{e_{C-N}}\left(A_{nm}^{C-N}\right)_3 + \overline{e_{C-Cl}}\left(A_{nm}^{C-Cl}\right)_3 + \overline{e_{C-S}}\left(A_{nm}^{C-S}\right)_3 + \overline{e_{O-C}}\left(A_{nm}^{O-C}\right)_3 + \overline{e_{O-O}}\left(A_{nm}^{O-O}\right)_3 + \overline{e_{O-N}}\left(A_{nm}^{O-N}\right)_3 \\
& + \overline{e_{O-Cl}}\left(A_{nm}^{O-Cl}\right)_3 + \overline{e_{O-S}}\left(A_{nm}^{O-S}\right)_3 + \overline{e_{N-C}}\left(A_{nm}^{N-C}\right)_3 + \overline{e_{N-O}}\left(A_{nm}^{N-O}\right)_3 + \overline{e_{N-N}}\left(A_{nm}^{N-N}\right)_3 + \overline{e_{N-Cl}}\left(A_{nm}^{N-Cl}\right)_3 + \overline{e_{N-S}}\left(A_{nm}^{N-S}\right)_3 + \overline{e_{Cl-C}}\left(A_{nm}^{Cl-C}\right)_3 \\
& \underbrace{+\overline{e_{Cl-O}}\left(A_{nm}^{Cl-O}\right)_3 + \overline{e_{Cl-N}}\left(A_{nm}^{Cl-N}\right)_3 + \overline{e_{Cl-Cl}}\left(A_{nm}^{Cl-Cl}\right)_3 + \overline{e_{Cl-S}}\left(A_{nm}^{Cl-S}\right)_3 + \overline{e_{S-C}}\left(A_{nm}^{S-C}\right)_3 + \overline{e_{S-O}}\left(A_{nm}^{S-O}\right)_3 + \overline{e_{S-N}}\left(A_{nm}^{S-N}\right)_3 + \overline{e_{S-Cl}}\left(A_{nm}^{S-Cl}\right)_3 + \overline{e_{S-S}}\left(A_{nm}^{S-S}\right)_3}_{3^{rd}-order\ interactions}
\end{aligned}
\tag{3.8}
$$

Where the coefficients A_{mn}^{XY} and A_{mn}^{XY} are described by Equations (3.3)-(3.6). The use of all the parameters is not necessary for a particular group interaction (as it will be shown in example 3.1).

3.4.2 Relation between Group and CI-Interaction Parameters

It should be noted that in the expressions for a_{nm} (Equation (3.7)) and a_{mn} (Equation (3.8)) not all the terms are necessarily used for a single calculation. The following example is used to demonstrate that not all the terms of Equations (3.7) and (3.8) are always needed.

Example 3.1

One of the main concerns of a model developer is to generate models that are easy to use and implement for quick and reliable estimates. The number of parameters involved in specific models is important since the use of a small number of parameters will be translated in more simple and easy-to-use models. The derived Equations (3.7) and (3.8) to calculate missing GIPs for the Original UNIFAC (VLE) method are generalized expressions which include up to two hundred of parameters. However, for a specific calculation (e.g. one or a pair of GIPs) not all the parameters and coefficients are needed, which makes the calculation of the GIPs very simple, even for manual calculations. This is highlighted in example 3.1.

If, for example, an interaction between the main groups ACH (main group 3), and CCN (main group 19) is needed for the CI approach, the following values (taken from Appendix A) of CIs and stoichiometry are used:

Table 3.3 CI and Stoichiometric Values for the UNIFAC Groups ACH and CCN

	n_C	n_O	n_N	${}^0\chi^v$	${}^1\chi^v$	${}^2\chi^v$
ACH	1	0	0	0.577350	0.666666	0.192450
CCN	2	0	1	1.447213	0.473606	0.111803

It is clear that we have 2 different types of atoms and all the possible orders of connectivity indices playing a role in this calculation. Once these values are defined, Equation (3.7) is reduced to:

$$a_{ACH-CCN} = b_{C-C}\left(A_{ACH-CCN}^{C-C}\right)_0 + c_{C-C}\left(A_{ACH-CCN}^{C-C}\right)_1 + d_{C-C}\left(A_{ACH-CCN}^{C-C}\right)_2 + e_{C-C}\left(A_{ACH-CCN}^{C-C}\right)_3$$
$$+ b_{C-N}\left(A_{ACH-CCN}^{C-N}\right)_0 + c_{C-N}\left(A_{ACH-CCN}^{C-N}\right)_1 + d_{C-N}\left(A_{ACH-CCN}^{C-N}\right)_2 + e_{C-N}\left(A_{ACH-CCN}^{C-N}\right)_3$$

This expression can be further developed with explicit values of CIs and stoichiometry:

$$a_{ACH-CCN} = b_{C-C}\frac{n_C^{(ACH)}\,{}^v\chi_{(CCN)}^0 - n_C^{(CCN)}\,{}^v\chi_{(ACH)}^0}{{}^v\chi_{(CCN)}^0\,{}^v\chi_{(ACH)}^0} + c_{C-C}\frac{n_C^{(ACH)}\,{}^v\chi_{(CCN)}^1 - n_C^{(CCN)}\,{}^v\chi_{(ACH)}^0}{{}^v\chi_{(CCN)}^1\,{}^v\chi_{(ACH)}^0}$$
$$+ d_{C-C}\frac{n_C^{(ACH)}\,{}^v\chi_{(CCN)}^1 - n_C^{(CCN)}\,{}^v\chi_{(ACH)}^1}{{}^v\chi_{(CCN)}^1\,{}^v\chi_{(ACH)}^1} + e_{C-C}\frac{n_C^{(ACH)}\,{}^v\chi_{(CCN)}^2 - n_C^{(CCN)}\,{}^v\chi_{(ACH)}^0}{{}^v\chi_{(CCN)}^2\,{}^v\chi_{(ACH)}^0}$$

$$+\ b_{C-N}\frac{n_C^{(ACH)}\,{}^{\nu}\chi^0_{(CCN)}-n_N^{(CCN)}\,{}^{\nu}\chi^0_{(ACH)}}{{}^{\nu}\chi^0_{(CCN)}\,{}^{\nu}\chi^0_{(ACH)}}+c_{C-N}\frac{n_C^{(ACH)}\,{}^{\nu}\chi^1_{(CCN)}-n_N^{(CCN)}\,{}^{\nu}\chi^0_{(ACH)}}{{}^{\nu}\chi^1_{(CCN)}\,{}^{\nu}\chi^0_{(ACH)}}$$
$$+d_{C-N}\frac{n_C^{(ACH)}\chi_1^{(CCN)}-n_N^{(CCN)}\,{}^{\nu}\chi^1_{(ACH)}}{\chi_1^{(CCN)}\,{}^{\nu}\chi^1_{(ACH)}}+e_{C-N}\frac{n_C^{(ACH)}\,{}^{\nu}\chi^2_{(CCN)}-n_N^{(CCN)}\,{}^{\nu}\chi^0_{(ACH)}}{{}^{\nu}\chi^2_{(CCN)}\,{}^{\nu}\chi^0_{(ACH)}}$$

The number of terms that are used for this calculation has been reduced from 100 terms in the generalized expression in Equation (3.7) to 8 terms, as can be seen above. No Cl and S atoms are involved in the calculation which reduces dramatically the number of terms used; the same applies when both of the UNIFAC groups are sufficiently small so that they can be described just with zeroth and first order connectivity indices. Even when it is necessary to use all the terms given in Equation (3.7), it does not represent a computational effort for the calculation, since it can be done in miliseconds using any conventional PC or laptop. The computational efforts appear when regressing the AIPs; this topic will be discussed in details in the following chapters.

3.4.3 Application Range

The atoms that the UNIFAC-CI model can handle are: C, O, N, Cl, and S (taking H implicitly into account). With the AIPs involving these five atoms, it might be possible to cover 330 out of 1306 gaps (25.26% of the gaps) that exist in the UNIFAC parameter table[7]. Table 3.4 summarizes the statistics related to the GIPs that can be filled in through the UNIFAC-CI method.

Table 3.4 Statistics when filling the orig. UNIFAC[71] parameter table

Filling of GIPs	Number of GIPs	Percentage of the Parameter Table	Percentage of the Missing GIPs
Total number of GIPs	2450	-	-
GIPs filled by Hansen et al.[71]	1144	46.69 %	-
GIPs missing in Hansen et al.[71] parameter table	1306	53.30 %	-
GIPs that can be potentially generated using UNIFAC-CI (C, O, N, Cl, and S atoms)	330	13.45 %	25.26 %

The rest of the gaps are divided into molecular groups for which there are no AIPs available and groups containing atoms other than C, O, N, Cl, and S. The main reason for not including these groups is that there are not enough experimental data (a minimum amount) available, and that certain atoms (C, O, N, Cl, and S atoms) were given priority in terms of computational complexity, computational time, and impact the have when filling the UNIFAC parameter table. However, the implementation of those missing functional and molecular groups is possible and should be considered for future work.

3.5 UNIFAC-CI FOR LLE

3.5.1 Reference Model (Original UNIFAC-VLE)

The work of Magnusson[72] has been taken into account for extension of the UNIFAC-CI method to LLE predictions. For this purpose, the original UNIFAC-LLE parameter table (a high number of gaps in the parameter table still exists) has been considered. In his work, Magnusson attempted to overcome the difficulties found in the original work of Fredenslund et al.[55] when dealing with LLE systems by trying to: '*establish a UNIFAC group contribution parameter table especially suited for the prediction of LLE*'. As it can be seen in Figure 3.3, the number of gaps in the parameter table is approximately 60%.

When using the GIPs parameters from the work of Fredenslund[55] and Skjold-Jørgensen[77], Magnusson[72] found that the agreement between experimental data and the predictions was mainly qualitative. One reason for this is that the data system used for regression of the GIPs was based on VLE data with temperatures far from those where LLE exists. Using simultaneously VLE and LLE data to generate the GIPs, brings the issue of 'average' parameters which pose a trade off between: a) using VLE and LLE data, that predict moderately well both VLE and LLE; and b) using just LLE data, where there is a loss of generality contrasted by good LLE predictions.

Finally, Magnusson decided to create a tailor-made UNIFAC group interaction parameter table for LLE, a decision well supported by the encouraging results obtained later when comparing the predictions, made at that time, between UNIFAC-LLE, UNIFAC-VLE and UNIQUAC (with individually regressed parameters).

The UNIFAC-LLE parameter table is composed of 32 main groups including: hydrocarbons, alcohols, acids, aromatics, ketones, water, aldehydes, nitrogen-containing compounds, halides, and sulfur-containing compouds. In the present work, C, O, H, and N related groups are implemented in the GC^{Plus} methodology. The current version of the parameter table proposed by Magnusson[72] in 1980 is highlighted in Figure 3.3, together with the GIPs that may be generated using CIs.

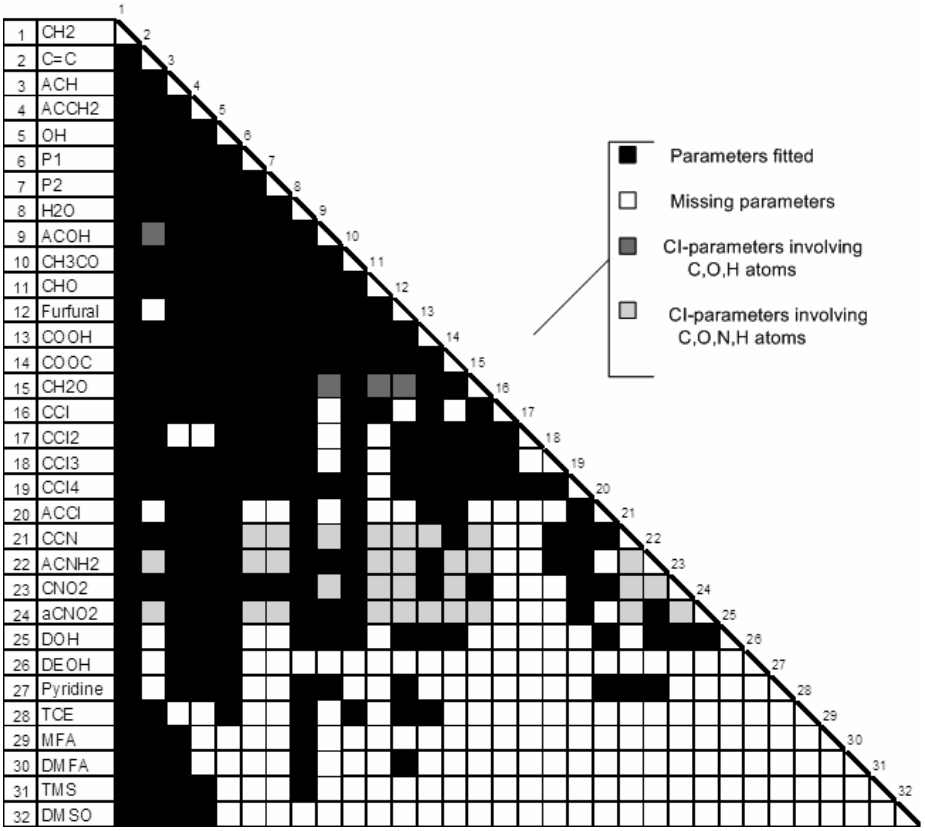


Figure 3.3 UNIFAC-LLE group interaction parameter matrix[72] + Group Interaction Parameters Generated Through Connectivity Indices

The values for the a_{mn} and a_{nm} parameters in original UNIFAC-LLE (see Equation (2.15)) can be calculated using the UNIFAC-CI (LLE) model as follows:

$$a_{mn} = \underbrace{\begin{aligned} &b_{C-C}\left(A_{mn}^{C-C}\right)_0 + b_{C-O}\left(A_{mn}^{C-O}\right)_0 + b_{C-N}\left(A_{mn}^{C-N}\right)_0 + b_{O-C}\left(A_{mn}^{O-C}\right)_0 \\ &+b_{O-O}\left(A_{mn}^{O-O}\right)_0 + b_{O-N}\left(A_{mn}^{O-N}\right)_0 + b_{N-C}\left(A_{mn}^{N-C}\right)_0 + b_{N-O}\left(A_{mn}^{N-O}\right)_0 + b_{N-N}\left(A_{mn}^{N-N}\right)_0 \end{aligned}}_{zeroth-order\ \text{interactions}}$$

$$\underbrace{\begin{aligned} &+c_{C-C}\left(A_{mn}^{C-C}\right)_1 + c_{C-O}\left(A_{mn}^{C-O}\right)_1 + c_{C-N}\left(A_{mn}^{C-N}\right)_1 + c_{O-C}\left(A_{mn}^{O-C}\right)_1 \\ &+c_{O-O}\left(A_{mn}^{O-O}\right)_1 + c_{O-N}\left(A_{mn}^{O-N}\right)_1 + c_{N-C}\left(A_{mn}^{N-C}\right)_1 + c_{N-O}\left(A_{mn}^{N-O}\right)_1 + c_{N-N}\left(A_{mn}^{N-N}\right)_1 \end{aligned}}_{1^{st}\text{-}order\ \text{interactions}}$$

$$\underbrace{\begin{aligned} &+d_{C-C}\left(A_{mn}^{C-C}\right)_2 + d_{C-O}\left(A_{mn}^{C-O}\right)_2 + d_{C-N}\left(A_{mn}^{C-N}\right)_2 + d_{O-C}\left(A_{mn}^{O-C}\right)_2 \\ &+d_{O-O}\left(A_{mn}^{O-O}\right)_2 + d_{O-N}\left(A_{mn}^{O-N}\right)_2 + d_{N-C}\left(A_{mn}^{N-C}\right)_2 + d_{N-O}\left(A_{mn}^{N-O}\right)_2 + d_{N-N}\left(A_{mn}^{N-N}\right)_2 \end{aligned}}_{2^{nd}\text{-}order\ \text{interactions}}$$

$$+e_{C-C}\left(A_{mn}^{C-C}\right)_3+e_{C-O}\left(A_{mn}^{C-O}\right)_3+e_{C-S}\left(A_{mn}^{C-S}\right)_3+e_{O-C}\left(A_{mn}^{O-C}\right)_3$$
$$\underbrace{+e_{O-O}\left(A_{mn}^{O-O}\right)_3+e_{O-N}\left(A_{mn}^{O-N}\right)_3+e_{N-C}\left(A_{mn}^{N-C}\right)_3+e_{N-O}\left(A_{mn}^{N-O}\right)_3+e_{N-N}\left(A_{mn}^{N-N}\right)_3}_{3^{rd}-order\ interactions} \quad (3.9)$$

and:

$$a_{nm}=\underbrace{\overline{b_{C-C}}\left(A_{nm}^{C-C}\right)_0+\overline{b_{C-O}}\left(A_{nm}^{C-O}\right)_0+\overline{b_{C-N}}\left(A_{nm}^{C-N}\right)_0+\overline{b_{O-C}}\left(A_{nm}^{O-C}\right)_0+\overline{b_{O-O}}\left(A_{nm}^{O-O}\right)_0+\overline{b_{O-N}}\left(A_{nm}^{O-N}\right)_0+\overline{b_{N-C}}\left(A_{nm}^{N-C}\right)_0+\overline{b_{N-O}}\left(A_{nm}^{N-O}\right)_0+\overline{b_{N-N}}\left(A_{nm}^{N-N}\right)_0}_{zeroth-order\ interactions}$$

$$\underbrace{+\overline{c_{C-C}}\left(A_{nm}^{C-C}\right)_1+\overline{c_{C-O}}\left(A_{nm}^{C-O}\right)_1+\overline{c_{C-N}}\left(A_{nm}^{C-N}\right)_1+\overline{c_{O-C}}\left(A_{nm}^{O-C}\right)_1+\overline{c_{O-O}}\left(A_{nm}^{O-O}\right)_1+\overline{c_{O-N}}\left(A_{nm}^{O-N}\right)_1+\overline{c_{N-C}}\left(A_{nm}^{N-C}\right)_1+\overline{c_{N-O}}\left(A_{nm}^{N-O}\right)_1+\overline{c_{N-N}}\left(A_{nm}^{N-N}\right)_1}_{1^{st}-order\ interactions}$$

$$\underbrace{+\overline{d_{C-C}}\left(A_{nm}^{C-C}\right)_2+\overline{d_{C-O}}\left(A_{nm}^{C-O}\right)_2+\overline{d_{C-N}}\left(A_{nm}^{C-N}\right)_2+\overline{d_{O-C}}\left(A_{nm}^{O-C}\right)_2+\overline{d_{O-O}}\left(A_{nm}^{O-O}\right)_2+\overline{d_{O-N}}\left(A_{nm}^{O-N}\right)_2+\overline{d_{N-C}}\left(A_{nm}^{N-C}\right)_2+\overline{d_{N-O}}\left(A_{nm}^{N-O}\right)_2+\overline{d_{N-N}}\left(A_{nm}^{N-N}\right)_2}_{2^{nd}-order\ interactions}$$

$$+\overline{e_{C-C}}\left(A_{nm}^{C-C}\right)_3+\overline{e_{C-O}}\left(A_{nm}^{C-O}\right)_3+\overline{e_{C-S}}\left(A_{nm}^{C-S}\right)_3+\overline{e_{O-C}}\left(A_{nm}^{O-C}\right)_3$$
$$\underbrace{+\overline{e_{O-O}}\left(A_{nm}^{O-O}\right)_3+\overline{e_{O-N}}\left(A_{nm}^{O-N}\right)_3+\overline{e_{N-C}}\left(A_{nm}^{N-C}\right)_3+\overline{e_{N-O}}\left(A_{nm}^{N-O}\right)_3+\overline{e_{N-N}}\left(A_{nm}^{N-N}\right)_3}_{3^{rd}-order\ interactions} \quad (3.10)$$

Where the coefficients A_{mn}^{XY} and A_{mn}^{XY} are described by Equations (3.3)-(3.6). Again, the use of all the parameters is not always necessary.

3.5.2 Relation between Group and CI –Interaction Parameters

The only difference when using the CIs and stoichiometry, with respect to the original UNIFAC-CI (VLE) model, is that, in this case, the identity of the groups might be different. The procedure for estimating the GIPs is however entirely the same. In the LLE parameter table, there are 'few' groups involving more than two different atoms (excluding hydrogen): for example, CNO_2 and $ACNO_2$. This means that there are 'a few' calculations involving more than a dozen of terms in the expressions derived for a_{mn} and a_{nm}, and even though an interaction between two groups of this type would be needed, that case does not pose any problem in terms of computational time since the calculations are performed within miliseconds.

An example of this kind of calculations is given below.

Example 3.2

In this example the way the GIPs are calculated for the UNIFAC-CI (LLE) groups is illustrated step-by-step. Here, the use of one UNIFAC group including up to three atoms and CIs of all orders is shown.

If the interaction between the main groups CH_2O (main group 13) and CNO_2 (main group 26) is needed when using UNIFAC-CI (LLE), the following values (taken from Appendix A) of CIs and stoichiometric values are used:

Table 3.5 CI and Stoichiometric Values for the UNIFAC-LLE Groups CH_2O and CNO_2

	n_C	n_O	n_N	${}^v\chi^0$	${}^v\chi^1$	${}^v\chi^2$
CH_2O	1	1	0	1.115355	0.288675	0.000000
CNO_2	1	2	1	2.224744	0.741581	0.401374

The condensed expression for the a_{mn} interaction (where m is CH_2O and n is CNO_2), including 24 terms, has the following form:

$$a_{CH_2O-CNO_2} = b_{C-C}\left(A^{C-C}_{CH_2O-CNO_2}\right)_0 + c_{C-C}\left(A^{C-C}_{CH_2O-CNO_2}\right)_1 + d_{C-C}\left(A^{C-C}_{CH_2O-CNO_2}\right)_2 + e_{C-C}\left(A^{C-C}_{CH_2O-CNO_2}\right)_3$$

$$b_{C-O}\left(A^{C-O}_{CH_2O-CNO_2}\right)_0 + c_{C-O}\left(A^{C-O}_{CH_2O-CNO_2}\right)_1 + d_{C-O}\left(A^{C-O}_{CH_2O-CNO_2}\right)_2 + e_{C-O}\left(A^{C-O}_{CH_2O-CNO_2}\right)_3$$

$$b_{C-N}\left(A^{C-N}_{CH_2O-CNO_2}\right)_0 + c_{C-N}\left(A^{C-N}_{CH_2O-CNO_2}\right)_1 + d_{C-N}\left(A^{C-N}_{CH_2O-CNO_2}\right)_2 + e_{C-N}\left(A^{C-N}_{CH_2O-CNO_2}\right)_3$$

$$+b_{O-C}\left(A^{O-C}_{CH_2O-CNO_2}\right)_0 + c_{O-C}\left(A^{O-C}_{CH_2O-CNO_2}\right)_1 + d_{O-C}\left(A^{O-C}_{CH_2O-CNO_2}\right)_2 + e_{O-C}\left(A^{O-C}_{CH_2O-CNO_2}\right)_3$$

$$+b_{O-O}\left(A^{O-O}_{CH_2O-CNO_2}\right)_0 + c_{O-O}\left(A^{O-O}_{CH_2O-CNO_2}\right)_1 + d_{O-O}\left(A^{O-O}_{CH_2O-CNO_2}\right)_2 + e_{O-O}\left(A^{O-O}_{CH_2O-CNO_2}\right)_3$$

$$+b_{O-N}\left(A^{O-N}_{CH_2O-CNO2}\right)_0 + c_{O-N}\left(A^{O-N}_{CH_2O-CNO2}\right)_1 + d_{O-N}\left(A^{O-N}_{CH_2O-CNO2}\right)_2 + e_{O-N}\left(A^{O-N}_{CH_2O-CNO2}\right)_3$$

The same approach applies for a_{nm}, including 18 terms with the mirror AIPs, for example, $\overline{b_{C-C}}$, $\overline{b_{C-O}}$, $\overline{b_{C-N}}$, etc. The relatively large size of the CNO_2 group results in accounting for CIs of all the orders (0^{th}, 1^{st}, and 2^{nd}) and atoms of all the available types (C, O, and N) for the calculation of the a_{mn} parameter in UNIFAC-CI LLE model.

3.5.3 Application Range

Just as for the UNIFAC-VLE parameter tables, (approximately 50%) there are group interaction parameters missing in the UNIFAC-LLE table. The exact figures are shown in Table 3.6, where a summary of the statistics regarding the number of GIPs calculated, missing, and generated through CIs, is given. The total number of parameters in the UNIFAC-LLE parameter table is 992, out of which 420 were estimated by Magnusson[72].

The scope of this work for UNIFAC-LLE has been directed towards including groups with C, O, and N atoms. So far, nitrogen-containing molecular groups (for example, pyridine) have been excluded, but they are considered for future work. The possible GIPs that can then be generated represent 70 GIPs out of 572 are missing (approximately 7.05% of the entire table, see Figure 3.3). Even though chlorine and sulfur atoms are missing (due to the lack of a minimum amount of experimental data),

the potential number of GIPs that can be filled in is still important, in addition to the fact that compounds including nitrogen are popular in LLE calculations.

It should be pointed out that no other documented attempt to extensively fill the Original UNIFAC-LLE parameter table is known at the moment. There are individual works where certain GIPs of particular interest have been regressed by matching experimental data[78, 79, 80], but these works had no intention of filling the entire UNIFAC parameter table. So, this means that the same gaps in the parameter table developed by Magnusson[72] remain until now.

Table 3.6 Statistics when filling the orig. UNIFAC-LLE[72] parameter table

Situation	Number of GIPs	Percentage of the Parameter Table	Percentage of the Missing GIPs
Total number of GIPs	992	-	-
GIPs filled by Magnusson et al.	420	42.33 %	-
GIPs missing in Magnusson´s parameter table	572	57.66 %	-
GIPs that can be potentially generated using UNIFAC-CI (C, O, and N atoms)	70	7.05 %	12.23 %

3.6 UNIFAC-CI (DORTMUND)

3.6.1 Reference Model (Modified UNIFAC-Dortmund)

The reference model used in this work is that of Gmehling et al[73]. The reason is that it is the most accessible publication when it comes to the availability of the group interaction parameters together with a clear explanation of the model, plus the fact that recent parameter revisions do not give the parameter values for all the newly determined GIPs (valuable for comparison purposes) and the subgroup representations of the UNIFAC main groups (in order to reproduce the same scheme when the Modified UNIFAC (Dortmund) and the CI-model are compared).

The reference Modified UNIFAC (Dortmund) model contains 45 main groups and their corresponding interaction parameters. As it can be seen in Figure 3.5, there are numerous gaps in the parameter table. Some of the main differences between Modified UNIFAC (Dortmund) and original UNIFAC rely on the fact that in the combinatorial term of Modified UNIFAC (Dortmund), the parameter V_i from Equation (3.11), has been empirically changed to improve the description of very asymmetric molecules:

$$V_i' = \frac{r_i^{3/4}}{\sum_j x_j r_j^{3/4}} \tag{3.11}$$

In addition, three sets of temperature dependent parameters have been included to allow a better description of the non-ideality as a function of temperature:

$$\Psi_{nm} = \exp\left(-\frac{a_{nm} + b_{nm}T + c_{nm}T^2}{T}\right) \tag{3.12}$$

van der Waals volume and surface parameters (R_k and Q_k) of the structural UNIFAC groups were not calculated using the *Bondi*[56] rules but optimized together with the GIPs parameters using experimental data.

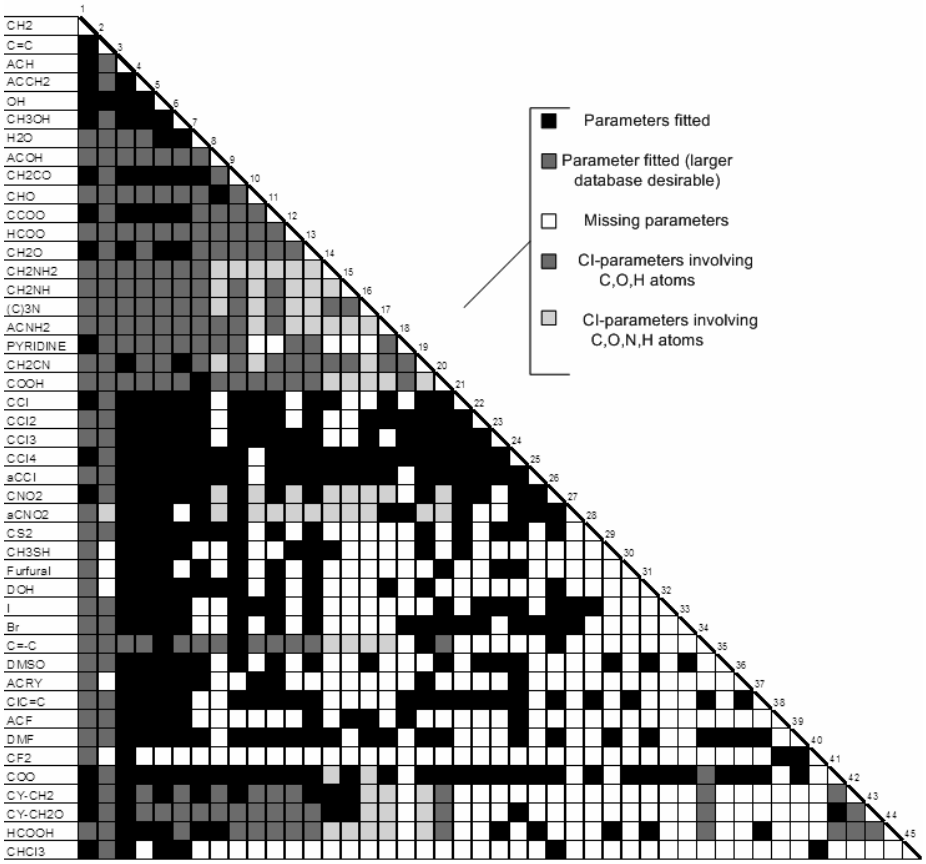


Figure 3.5 Modified UNIFAC-CI (Dortmund) group interaction parameter matrix[73] *+ Group Interaction Parameters generated through Connectivity Indices.*

3.6.2 Relation between Group and CI –Interaction Parameters

The values for the a_{nm}, b_{nm}, and c_{nm} (and their mirror) parameters in Modified UNIFAC Dortmund (see Equation (3.12)) can be calculated using the UNIFAC-CI (Dortmund) model as follows:

$$a_{mn} = \underbrace{\begin{aligned} & b^1_{C-C}\left(A^{C-C}_{mn}\right)_0 + b^1_{C-O}\left(A^{C-O}_{mn}\right)_0 + b^1_{C-N}\left(A^{C-N}_{mn}\right)_0 + b^1_{O-C}\left(A^{O-C}_{mn}\right)_0 \\ & + b^1_{O-O}\left(A^{O-O}_{mn}\right)_0 + b^1_{O-N}\left(A^{O-N}_{mn}\right)_0 + b^1_{N-C}\left(A^{N-C}_{mn}\right)_0 + b^1_{N-O}\left(A^{N-O}_{mn}\right)_0 + b^1_{N-N}\left(A^{N-N}_{mn}\right)_0 \end{aligned}}_{zeroth-order\ \text{interactions}}$$

$$
\begin{aligned}
&+c^1_{C-C}\left(A^{C-C}_{mn}\right)_1+c^1_{C-O}\left(A^{C-O}_{mn}\right)_1+c^1_{C-N}\left(A^{C-N}_{mn}\right)_1+c^1_{O-C}\left(A^{O-C}_{mn}\right)_1\\
&\underbrace{+c^1_{O-O}\left(A^{O-O}_{mn}\right)_1+c^1_{O-N}\left(A^{O-N}_{mn}\right)_1+c^1_{N-C}\left(A^{N-C}_{mn}\right)_1+c^1_{N-O}\left(A^{N-O}_{mn}\right)_1+c^1_{N-N}\left(A^{N-N}_{mn}\right)_1}_{1^{st}\text{-}order\ \text{interactions}}\\
&+d^1_{C-C}\left(A^{C-C}_{mn}\right)_2+d^1_{C-O}\left(A^{C-O}_{mn}\right)_2+d^1_{C-N}\left(A^{C-N}_{mn}\right)_2+d^1_{O-C}\left(A^{O-C}_{mn}\right)_2\\
&\underbrace{+d^1_{O-O}\left(A^{O-O}_{mn}\right)_2+d^1_{O-N}\left(A^{O-N}_{mn}\right)_2+d^1_{N-C}\left(A^{N-C}_{mn}\right)_2+d^1_{N-O}\left(A^{N-O}_{mn}\right)_2+d^1_{N-N}\left(A^{N-N}_{mn}\right)_2}_{2^{nd}\text{-}order\ \text{interactions}}\\
&+e^1_{C-C}\left(A^{C-C}_{mn}\right)_3+e^1_{C-O}\left(A^{C-O}_{mn}\right)_3+e^1_{C-S}\left(A^{C-S}_{mn}\right)_3+e^1_{O-C}\left(A^{O-C}_{mn}\right)_3\\
&\underbrace{+e^1_{O-O}\left(A^{O-O}_{mn}\right)_3+e^1_{O-N}\left(A^{O-N}_{mn}\right)_3+e^1_{N-C}\left(A^{N-C}_{mn}\right)_3+e^1_{N-O}\left(A^{N-O}_{mn}\right)_3+e^1_{N-N}\left(A^{N-N}_{mn}\right)_3}_{3^{rd}-order\ \text{interactions}}
\end{aligned}
\tag{3.13}
$$

$$
\begin{aligned}
b_{mn}=\;&b^2_{C-C}\left(A^{C-C}_{mn}\right)_0+b^2_{C-O}\left(A^{C-O}_{mn}\right)_0+b^2_{C-N}\left(A^{C-N}_{mn}\right)_0+b^2_{O-C}\left(A^{O-C}_{mn}\right)_0\\
&\underbrace{+b^2_{O-O}\left(A^{O-O}_{mn}\right)_0+b^2_{O-N}\left(A^{O-N}_{mn}\right)_0+b^2_{N-C}\left(A^{N-C}_{mn}\right)_0+b^2_{N-O}\left(A^{N-O}_{mn}\right)_0+b^2_{N-N}\left(A^{N-N}_{mn}\right)_0}_{zeroth-order\ \text{interactions}}\\
&+c^2_{C-C}\left(A^{C-C}_{mn}\right)_1+c^2_{C-O}\left(A^{C-O}_{mn}\right)_1+c^2_{C-N}\left(A^{C-N}_{mn}\right)_1+c^2_{O-C}\left(A^{O-C}_{mn}\right)_1\\
&\underbrace{+c^2_{O-O}\left(A^{O-O}_{mn}\right)_1+c^2_{O-N}\left(A^{O-N}_{mn}\right)_1+c^2_{N-C}\left(A^{N-C}_{mn}\right)_1+c^2_{N-O}\left(A^{N-O}_{mn}\right)_1+c^2_{N-N}\left(A^{N-N}_{mn}\right)_1}_{1^{st}\text{-}order\ \text{interactions}}\\
&+d^2_{C-C}\left(A^{C-C}_{mn}\right)_2+d^2_{C-O}\left(A^{C-O}_{mn}\right)_2+d^2_{C-N}\left(A^{C-N}_{mn}\right)_2+d^2_{O-C}\left(A^{O-C}_{mn}\right)_2\\
&\underbrace{+d^2_{O-O}\left(A^{O-O}_{mn}\right)_2+d^2_{O-N}\left(A^{O-N}_{mn}\right)_2+d^2_{N-C}\left(A^{N-C}_{mn}\right)_2+d^2_{N-O}\left(A^{N-O}_{mn}\right)_2+d^2_{N-N}\left(A^{N-N}_{mn}\right)_2}_{2^{nd}\text{-}order\ \text{interactions}}\\
&+e^2_{C-C}\left(A^{C-C}_{mn}\right)_3+e^2_{C-O}\left(A^{C-O}_{mn}\right)_3+e^2_{C-S}\left(A^{C-S}_{mn}\right)_3+e^2_{O-C}\left(A^{O-C}_{mn}\right)_3\\
&\underbrace{+e^2_{O-O}\left(A^{O-O}_{mn}\right)_3+e^2_{O-N}\left(A^{O-N}_{mn}\right)_3+e^2_{N-C}\left(A^{N-C}_{mn}\right)_3+e^2_{N-O}\left(A^{N-O}_{mn}\right)_3+e^2_{N-N}\left(A^{N-N}_{mn}\right)_3}_{3^{rd}-order\ \text{interactions}}
\end{aligned}
\tag{3.14}
$$

$$
\begin{aligned}
c_{mn}=\;&b^3_{C-C}\left(A^{C-C}_{mn}\right)_0+b^3_{C-O}\left(A^{C-O}_{mn}\right)_0+b^3_{C-N}\left(A^{C-N}_{mn}\right)_0+b^3_{O-C}\left(A^{O-C}_{mn}\right)_0\\
&\underbrace{+b^3_{O-O}\left(A^{O-O}_{mn}\right)_0+b^3_{O-N}\left(A^{O-N}_{mn}\right)_0+b^3_{N-C}\left(A^{N-C}_{mn}\right)_0+b^3_{N-O}\left(A^{N-O}_{mn}\right)_0+b^3_{N-N}\left(A^{N-N}_{mn}\right)_0}_{zeroth-order\ \text{interactions}}\\
&+c^3_{C-C}\left(A^{C-C}_{mn}\right)_1+c^3_{C-O}\left(A^{C-O}_{mn}\right)_1+c^3_{C-N}\left(A^{C-N}_{mn}\right)_1+c^3_{O-C}\left(A^{O-C}_{mn}\right)_1\\
&\underbrace{+c^3_{O-O}\left(A^{O-O}_{mn}\right)_1+c^3_{O-N}\left(A^{O-N}_{mn}\right)_1+c^3_{N-C}\left(A^{N-C}_{mn}\right)_1+c^3_{N-O}\left(A^{N-O}_{mn}\right)_1+c^3_{N-N}\left(A^{N-N}_{mn}\right)_1}_{1^{st}\text{-}order\ \text{interactions}}\\
&+d^3_{C-C}\left(A^{C-C}_{mn}\right)_2+d^3_{C-O}\left(A^{C-O}_{mn}\right)_2+d^3_{C-N}\left(A^{C-N}_{mn}\right)_2+d^3_{O-C}\left(A^{O-C}_{mn}\right)_2\\
&\underbrace{+d^3_{O-O}\left(A^{O-O}_{mn}\right)_2+d^3_{O-N}\left(A^{O-N}_{mn}\right)_2+d^3_{N-C}\left(A^{N-C}_{mn}\right)_2+d^3_{N-O}\left(A^{N-O}_{mn}\right)_2+d^3_{N-N}\left(A^{N-N}_{mn}\right)_2}_{2^{nd}\text{-}order\ \text{interactions}}
\end{aligned}
$$

$$
\begin{aligned}
&+e^{3}_{C-C}\left(A^{C-C}_{mn}\right)_3+e^{3}_{C-O}\left(A^{C-O}_{mn}\right)_3+e^{3}_{C-S}\left(A^{C-S}_{mn}\right)_3+e^{3}_{O-C}\left(A^{O-C}_{mn}\right)_3\\
&\underbrace{+e^{3}_{O-O}\left(A^{O-O}_{mn}\right)_3+e^{3}_{O-N}\left(A^{O-N}_{mn}\right)_3+e^{3}_{N-C}\left(A^{N-C}_{mn}\right)_3+e^{3}_{N-O}\left(A^{N-O}_{mn}\right)_3+e^{3}_{N-N}\left(A^{N-N}_{mn}\right)_3}_{3^{rd}-order\ interactions}
\end{aligned}
\tag{3.15}
$$

And for the mirror parameters:

$$
\begin{aligned}
a_{nm}=&\overline{b^{1}_{C-C}}\left(A^{C-C}_{nm}\right)_0+\overline{b^{1}_{C-O}}\left(A^{C-O}_{nm}\right)_0+\overline{b^{1}_{C-N}}\left(A^{C-N}_{nm}\right)_0+\overline{b^{1}_{O-C}}\left(A^{O-C}_{nm}\right)_0\\
&\underbrace{+\overline{b^{1}_{O-O}}\left(A^{O-O}_{nm}\right)_0+\overline{b^{1}_{O-N}}\left(A^{O-N}_{nm}\right)_0+\overline{b^{1}_{N-C}}\left(A^{N-C}_{nm}\right)_0+\overline{b^{1}_{N-O}}\left(A^{N-O}_{nm}\right)_0+\overline{b^{1}_{N-N}}\left(A^{N-N}_{nm}\right)_0}_{zeroth-order\ interactions}\\
&+\overline{c^{1}_{C-C}}\left(A^{C-C}_{nm}\right)_1+\overline{c^{1}_{C-O}}\left(A^{C-O}_{nm}\right)_1+\overline{c^{1}_{C-N}}\left(A^{C-N}_{nm}\right)_1+\overline{c^{1}_{O-C}}\left(A^{O-C}_{nm}\right)_1\\
&\underbrace{+\overline{c^{1}_{O-O}}\left(A^{O-O}_{nm}\right)_1+\overline{c^{1}_{O-N}}\left(A^{O-N}_{nm}\right)_1+\overline{c^{1}_{N-C}}\left(A^{N-C}_{nm}\right)_1+\overline{c^{1}_{N-O}}\left(A^{N-O}_{nm}\right)_1+\overline{c^{1}_{N-N}}\left(A^{N-N}_{nm}\right)_1}_{1^{st}-order\ interactions}\\
&+\overline{d^{1}_{C-C}}\left(A^{C-C}_{nm}\right)_2+\overline{d^{1}_{C-O}}\left(A^{C-O}_{nm}\right)_2+\overline{d^{1}_{C-N}}\left(A^{C-N}_{nm}\right)_2+\overline{d^{1}_{O-C}}\left(A^{O-C}_{nm}\right)_2\\
&\underbrace{+\overline{d^{1}_{O-O}}\left(A^{O-O}_{nm}\right)_2+\overline{d^{1}_{O-N}}\left(A^{O-N}_{nm}\right)_2+\overline{d^{1}_{N-C}}\left(A^{N-C}_{nm}\right)_2+\overline{d^{1}_{N-O}}\left(A^{N-O}_{nm}\right)_2+\overline{d^{1}_{N-N}}\left(A^{N-N}_{nm}\right)_2}_{2^{nd}-order\ interactions}\\
&+\overline{e^{1}_{C-C}}\left(A^{C-C}_{nm}\right)_3+\overline{e^{1}_{C-O}}\left(A^{C-O}_{nm}\right)_3+\overline{e^{1}_{C-S}}\left(A^{C-S}_{nm}\right)_3+\overline{e^{1}_{O-C}}\left(A^{O-C}_{nm}\right)_3\\
&\underbrace{+\overline{e^{1}_{O-O}}\left(A^{O-O}_{nm}\right)_3+\overline{e^{1}_{O-N}}\left(A^{O-N}_{nm}\right)_3+\overline{e^{1}_{N-C}}\left(A^{N-C}_{nm}\right)_3+\overline{e^{1}_{N-O}}\left(A^{N-O}_{nm}\right)_3+\overline{e^{1}_{N-N}}\left(A^{N-N}_{nm}\right)_3}_{3^{rd}-order\ interactions}
\end{aligned}
\tag{3.16}
$$

$$
\begin{aligned}
b_{nm}=&\overline{b^{2}_{C-C}}\left(A^{C-C}_{nm}\right)_0+\overline{b^{2}_{C-O}}\left(A^{C-O}_{nm}\right)_0+\overline{b^{2}_{C-N}}\left(A^{C-N}_{nm}\right)_0+\overline{b^{2}_{O-C}}\left(A^{O-C}_{nm}\right)_0\\
&\underbrace{+\overline{b^{2}_{O-O}}\left(A^{O-O}_{nm}\right)_0+\overline{b^{2}_{O-N}}\left(A^{O-N}_{nm}\right)_0+\overline{b^{2}_{N-C}}\left(A^{N-C}_{nm}\right)_0+\overline{b^{2}_{N-O}}\left(A^{N-O}_{nm}\right)_0+\overline{b^{2}_{N-N}}\left(A^{N-N}_{nm}\right)_0}_{zeroth-order\ interactions}\\
&+\overline{c^{2}_{C-C}}\left(A^{C-C}_{nm}\right)_1+\overline{c^{2}_{C-O}}\left(A^{C-O}_{nm}\right)_1+\overline{c^{2}_{C-N}}\left(A^{C-N}_{nm}\right)_1+\overline{c^{2}_{O-C}}\left(A^{O-C}_{nm}\right)_1\\
&\underbrace{+\overline{c^{2}_{O-O}}\left(A^{O-O}_{nm}\right)_1+\overline{c^{2}_{O-N}}\left(A^{O-N}_{nm}\right)_1+\overline{c^{2}_{N-C}}\left(A^{N-C}_{nm}\right)_1+\overline{c^{2}_{N-O}}\left(A^{N-O}_{nm}\right)_1+\overline{c^{2}_{N-N}}\left(A^{N-N}_{nm}\right)_1}_{1^{st}-order\ interactions}\\
&+\overline{d^{2}_{C-C}}\left(A^{C-C}_{nm}\right)_2+\overline{d^{2}_{C-O}}\left(A^{C-O}_{nm}\right)_2+\overline{d^{2}_{C-N}}\left(A^{C-N}_{nm}\right)_2+\overline{d^{2}_{O-C}}\left(A^{O-C}_{nm}\right)_2\\
&\underbrace{+\overline{d^{2}_{O-O}}\left(A^{O-O}_{nm}\right)_2+\overline{d^{2}_{O-N}}\left(A^{O-N}_{nm}\right)_2+\overline{d^{2}_{N-C}}\left(A^{N-C}_{nm}\right)_2+\overline{d^{2}_{N-O}}\left(A^{N-O}_{nm}\right)_2+\overline{d^{2}_{N-N}}\left(A^{N-N}_{nm}\right)_2}_{2^{nd}-order\ interactions}\\
&+\overline{e^{2}_{C-C}}\left(A^{C-C}_{nm}\right)_3+\overline{e^{2}_{C-O}}\left(A^{C-O}_{nm}\right)_3+\overline{e^{2}_{C-S}}\left(A^{C-S}_{nm}\right)_3+\overline{e^{2}_{O-C}}\left(A^{O-C}_{nm}\right)_3\\
&\underbrace{+\overline{e^{2}_{O-O}}\left(A^{O-O}_{nm}\right)_3+\overline{e^{2}_{O-N}}\left(A^{O-N}_{nm}\right)_3+\overline{e^{2}_{N-C}}\left(A^{N-C}_{nm}\right)_3+\overline{e^{2}_{N-O}}\left(A^{N-O}_{nm}\right)_3+\overline{e^{2}_{N-N}}\left(A^{N-N}_{nm}\right)_3}_{3^{rd}-order\ interactions}
\end{aligned}
\tag{3.17}
$$

$$
\begin{aligned}
c_{nm} = {} & \overline{b^3_{C-C}}\left(A^{C-C}_{nm}\right)_0 + \overline{b^3_{C-O}}\left(A^{C-O}_{nm}\right)_0 + \overline{b^3_{C-N}}\left(A^{C-N}_{nm}\right)_0 + \overline{b^3_{O-C}}\left(A^{O-C}_{nm}\right)_0 \\
& \underbrace{+\overline{b^3_{O-O}}\left(A^{O-O}_{nm}\right)_0 + \overline{b^3_{O-N}}\left(A^{O-N}_{nm}\right)_0 + \overline{b^3_{N-C}}\left(A^{N-C}_{nm}\right)_0 + \overline{b^3_{N-O}}\left(A^{N-O}_{nm}\right)_0 + \overline{b^3_{N-N}}\left(A^{N-N}_{nm}\right)_0}_{zeroth-order\ \text{interactions}} \\
& + \overline{c^3_{C-C}}\left(A^{C-C}_{nm}\right)_1 + \overline{c^3_{C-O}}\left(A^{C-O}_{nm}\right)_1 + \overline{c^3_{C-N}}\left(A^{C-N}_{nm}\right)_1 + \overline{c^3_{O-C}}\left(A^{O-C}_{nm}\right)_1 \\
& \underbrace{+\overline{c^3_{O-O}}\left(A^{O-O}_{nm}\right)_1 + \overline{c^3_{O-N}}\left(A^{O-N}_{nm}\right)_1 + \overline{c^3_{N-C}}\left(A^{N-C}_{nm}\right)_1 + \overline{c^3_{N-O}}\left(A^{N-O}_{nm}\right)_1 + \overline{c^3_{N-N}}\left(A^{N-N}_{nm}\right)_1}_{1^{st}\text{-}order\ \text{interactions}} \\
& + \overline{d^3_{C-C}}\left(A^{C-C}_{nm}\right)_2 + \overline{d^3_{C-O}}\left(A^{C-O}_{nm}\right)_2 + \overline{d^3_{C-N}}\left(A^{C-N}_{nm}\right)_2 + \overline{d^3_{O-C}}\left(A^{O-C}_{nm}\right)_2 \\
& \underbrace{+\overline{d^3_{O-O}}\left(A^{O-O}_{nm}\right)_2 + \overline{d^3_{O-N}}\left(A^{O-N}_{nm}\right)_2 + \overline{d^3_{N-C}}\left(A^{N-C}_{nm}\right)_2 + \overline{d^3_{N-O}}\left(A^{N-O}_{nm}\right)_2 + \overline{d^3_{N-N}}\left(A^{N-N}_{nm}\right)_2}_{2^{nd}\text{-}order\ \text{interactions}} \\
& + \overline{e^3_{C-C}}\left(A^{C-C}_{nm}\right)_3 + \overline{e^3_{C-O}}\left(A^{C-O}_{nm}\right)_3 + \overline{e^3_{C-S}}\left(A^{C-S}_{nm}\right)_3 + \overline{e^3_{O-C}}\left(A^{O-C}_{nm}\right)_3 \\
& \underbrace{+\overline{e^3_{O-O}}\left(A^{O-O}_{nm}\right)_3 + \overline{e^3_{O-N}}\left(A^{O-N}_{nm}\right)_3 + \overline{e^3_{N-C}}\left(A^{N-C}_{nm}\right)_3 + \overline{e^3_{N-O}}\left(A^{N-O}_{nm}\right)_3 + \overline{e^3_{N-N}}\left(A^{N-N}_{nm}\right)_3}_{3^{rd}-order\ \text{interactions}}
\end{aligned}
\tag{3.18}
$$

Since now there are 3 sets of GIPs and subsequently three subsets of AIPs, the superscripts 1, 2, and 3 relate the GIPs to the corresponding AIPs: a_{mn}, b_{nm}, and c_{nm} respectively.

The coefficients A^{XY}_{mn} and A^{XY}_{nm} are calculated from Equations (3.3) - (3.6). The use of all the parameters is again not always necessary.

Originally, c_{nm} group interaction parameters (that follow a quadratic temperature dependence) were regressed using the experimental data for H^E (T) and $c_P{}^E$, in addition to VLE data[73]. In this work, the parameter c_{nm} has been set to zero because there were not enough experimental data for taking into account extreme sensitivity to temperature effects. The aim was also to make the optimization process easy to monitor and the correlation analysis as simple and fast as possible. Experimental data other than VLE (T) were not used. Adding c_{nm} parameter to the CI-model will be taken into account in future work.

Adding another set of parameters does not pose a computational problem when generating GIPs. It might however, slow the computational time when optimizing the AIPs (this issue will be addressed in the following chapters). For comparison purposes, this example illustrates the generation of GIPs for Modified UNIFAC-CI (Dortmund).

Example 3.3

The use of Equations (3.3) - (3.6) is explained in this example for a case where a group interaction is needed. The calculations are performed step by step to illustrate the simplicity of the implementation of a new GC model into the methodology.

If an interaction between the main groups ACH (main group 3) and CCOO (main group 11) is needed to be calculated through the the CI approach, the following values (taken from Appendix A) of CI and stoichiometry are used:

Table 3.7 CI and Stoichiometric Values for the UNIFAC Groups ACH and CCOO

	n_C	n_O	n_N	${}^{v}\chi^{0}$	${}^{v}\chi^{1}$	${}^{v}\chi^{2}$
ACH	1	0	0	0.577350	0.666666	0.192450
CCOO	2	2	0	1.816496	0.658248	0.287457

There are 2 different types of atoms and all the possible orders of connectivity indices are playing a role in this calculation. Once these values are defined, Equations (3.13) and (3.14) are reduced in the following way, for the a_{mn} and b_{nm} parameters (note that c_{nm} is not taken into account for this version of the CI-model), where n is ACH group and m is CCOO group:

$$a_{ACH-CCOO} = b^{1}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_0 + c^{1}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_1 + d^{1}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_2 + e^{1}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_3 + b^{1}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_0 + c^{1}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_1 + d^{1}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_2 + e^{1}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_3$$

$$b_{ACH-CCOO} = b^{2}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_0 + c^{2}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_1 + d^{2}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_2 + e^{2}{}_{C-C}\left(A^{C-C}_{ACH-CCOO}\right)_3 + b^{2}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_0 + c^{2}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_1 + d^{2}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_2 + e^{2}{}_{C-O}\left(A^{C-O}_{ACH-CCOO}\right)_3$$

Expanding the $\left(A^{X-Y}_{mn}\right)_i$ coefficient, both expressions obtain the following form:

$$a_{ACH-CCOO} = b^{1}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{0}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{0}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + c^{1}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{1}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{1}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + d^{1}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{1}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}}{{}^{v}\chi^{1}_{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}} + e^{1}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{2}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{2}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + b^{1}{}_{C-O}\frac{n_C^{(ACH)}\,{}^{v}\chi^{0}_{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{0}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + c^{1}{}_{C-O}\frac{n_C^{(ACH)}\,{}^{v}\chi^{1}_{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{1}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + d^{1}{}_{C-O}\frac{n_C^{(ACH)}\,\chi_1^{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}}{\chi_1^{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}} + e^{1}{}_{C-O}\frac{n_C^{(ACH)}\,{}^{v}\chi^{2}_{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{2}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}$$

$$b_{ACH-CCOO} = b^{2}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{0}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{0}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + c^{2}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{1}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{1}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + d^{2}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{1}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}}{{}^{v}\chi^{1}_{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}} + e^{2}{}_{C-C}\frac{n_C^{(ACH)}\,{}^{v}\chi^{2}_{(CCOO)} - n_C^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{2}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + b^{2}{}_{C-O}\frac{n_C^{(ACH)}\,{}^{v}\chi^{0}_{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{0}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + c^{2}{}_{C-O}\frac{n_C^{(ACH)}\,{}^{v}\chi^{1}_{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{1}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}} + d^{2}{}_{C-O}\frac{n_C^{(ACH)}\,\chi_1^{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}}{\chi_1^{(CCOO)}\,{}^{v}\chi^{1}_{(ACH)}} + e^{2}{}_{C-O}\frac{n_C^{(ACH)}\,{}^{v}\chi^{2}_{(CCOO)} - n_O^{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}{{}^{v}\chi^{2}_{(CCOO)}\,{}^{v}\chi^{0}_{(ACH)}}$$

One difference between a_{mn} and b_{nm} parameters is the identity of the AIPs used for their calculations ($b^1_{X\text{-}Y}$, $c^1_{X\text{-}Y}$, $d^1_{X\text{-}Y}$, $e^1_{X\text{-}Y}$, $b^2_{X\text{-}Y}$, $c^2_{X\text{-}Y}$, $d^2_{X\text{-}Y}$, $e^2_{X\text{-}Y}$). From the maximum of 240 AIPs that can be used for UNIFAC-CI (Dortmund) when using C and O atoms, the generation of the ACH-CCOO group interaction parameter uses only 16 AIPs.

3.6.3 Application Range

The UNIFAC-CI (Dortmund) model has been developed to include C, O, and N atoms, which means that any group interaction in the parameter table containing those atoms can be generated. This also applies for further developments in the parameter table (additions of new groups) and also creation of new groups.

Table 3.8, includes the statistics of the 'available' GIPs and the gaps in the parameter table. An important number of the gaps in the parameter table (228 out of 926) can be potentially generated using the CI-model.

Table 3.8 Statistics when filling in the Modified UNIFAC (Dortmund)[73] parameter table

Situation	Number of GIPs	Percentage of the Parameter Table	Percentage of the Missing GIPs
Total number of GIPs	1980	-	-
GIPs filled by Gmehling et al.[73]	1054	53 %	-
GIPs missing in Gmehling´s[73] parameter table	926	47 %	-
GIPs that can be potentially generated using UNIFAC-CI (C, O, and N atoms)	228	11.51 %	24.62 %

The recommended temperature range for UNIFAC-CI (Dortmund) model is between 273-400 K. Using the parameters outside this range of temperature should be avoided.

In principle, the generated GIPs can be applied for γ^{∞}, h^E, c_P^E and LLE (just like for Modified UNIFAC Dortmund). However, since no γ^{∞}, h^E, c_P^E and LLE experimental data were used to regress the AIPs, it is not possible to assess the quality of results when using CI-generated GIPs for this kind of thermodynamic values. The introduction of γ^{∞}, h^E, c_P^E data will be considered after a thorough analysis of the results that can be achieved using the current version of the CI-models. At this stage of the project, the analysis of correlation and prediction results are easy to monitor using just VLE data.

Chapter 4

Parameter Estimation

4.1 INTRODUCTION

A very important step when developing and updating semi-empirical models such as ASOG/UNIFAC is the estimation of the parameters involved since a correlation of experimental data is needed for reasonable predictions. Since the AIPs are necessary to back-calculate the GIPs as was discussed in Chapters 2 and 3, the estimation of AIPs is a fundamental issue that needs to be addressed. Once it is clear that a parameter estimation procedure is needed to estimate the AIPs, the question of whether to match the parameters to experimental data or to existing GIPs, should be considered. In this work, it has been chosen to match the AIPs to experimental data rather than to existing tables of GIPs due to the fact that the overall objective of the modeling is to match the corresponding (experimental measured) phase behaviour.

4.2 PARAMETER ESTIMATION TECHNIQUE

4.2.1 Minimization Technique

A least squares approach has been chosen in this work in order to regress the parameters. A discussion about the details of the minimization technique applied, and some provisions taken to asses the correctness of the computations are given in Appendix C.

4.2.2 Objective Functions

The objective functions are very important, since a good and meaningful selection can lead to shorter computational times and a better set of parameters for the phenomena that needs to be matched. The correct selection is particularly important. Different objective functions have been applied for each of the UNIFAC-CI models. In order to have a fair comparison of the UNIFAC-CI models with their reference UNIFAC models, the original objective functions have been preserved whenever possible.

UNIFAC-CI (VLE) for reference model Original UNIFAC-VLE

The parameter estimation for this work is based exclusively on P, x_i, y_i and P, x_i VLE data. Isothermal data have been used, because, in this case, it is not necessary to specify the vapor pressure or its dependence on temperature. VLE experimental data have been checked for thermodynamic consistency using the test of Van Ness[74]. Due to the differences in units and the magnitude of pressure values in the isothermal data sets, the following objective function was chosen, which is expressed as the average relative pressure quadratic deviation:

$$OF = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_{i-exp} - P_{i-calc}}{P_{i-exp}}\right)^2 \tag{4.1}$$

where N is the number of experimental data points used for the estimation.

For systems not needing association terms, the equilibrium pressure was calculated as follows:

$$P_{calc} = \sum_i x_i \gamma_i P_i^{sat} \left(POY_i\right) \tag{4.2}$$

where i is an index running over all species in the mixture, and POY_i is the Poynting correction.

For systems needing an association term, the bubble and dew point calculations using association constants based on the method of Hayden and O´Connell[81], were employed.

To ensure that the parameters did not move too far away from the previous step, and thus, to guarantee that the previously investigated groups/data sets would still be represented with comparable accuracy, a parameter regularization was performed by adding a second term in the objective function, as follows:

$$OF = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_{i-exp} - P_{i-calc}}{P_{i-exp}}\right)^2 + w_{reg}\sum_j \left(AIP_j - AIP_j^{IG}\right)^2 \tag{4.3}$$

where N is the number of experimental data points, AIP_j is the current value of the CI-interaction parameter j, AIP_j^{IG} its corresponding initial guess, and w_{reg} is a weighting value used to increase or decrease the influence of regularization in the optimization.

UNIFAC-CI (LLE) for reference model Original UNIFAC-LLE

The parameter estimation for this model was performed by using only the binary LLE data. When regressing parameters for LLE, there are usually two different approaches, in terms of the least squares principle:

1) The minimization of activity coefficients:

$$OF = \sum_{i=1}^{N}\left(a_i^I - a_i^{II}\right)^2 \tag{4.4}$$

where a_i is the activity for phase I or II, expressed by $a_i = x_i\gamma_i$; and N denotes the number of experimental data points used in the estimation.

2) The minimization of phase compositions in one phase:

$$OF = \sum_{i=1}^{I}\sum_{j=1}^{N}\left(x_{ij}^{cal} - x_{ij}^{exp}\right)^2 \tag{4.5}$$

where N denotes the number of experimental data points used in the estimation, and I denotes the number of phases involved in the calculation. Weight factors are sometimes used in this expression. Both Equations (4.4) and (4.5) are termed *absolute* objective functions, while there are other expressions (using the same quantities) termed *relative* objective functions. These are expressed as:

$$OF = \sum_{i=1}^{I}\sum_{j=1}^{N}\left(\ln x_{ij}^{cal} - \ln x_{ij}^{exp}\right)^2 \tag{4.6}$$

or

$$OF = \sum_{i=1}^{I}\sum_{j=1}^{N}\left(\frac{x_{ij}^{cal} - x_{ij}^{exp}}{x_{ij}^{cal} + x_{ij}^{exp}}\right)^2 \tag{4.7}$$

Another important criterion for the achievement of phase equilibria is the equality of activities (isoactivity) between the phases. The isoactivity criterion is easier to implement than an objective function based on compositions, and does not require good initial guesses for convergence. This criterion, however, does not match the purpose of this work which is related to the minimization of the differences between experimental and calculated composition of a specific liquid phase.

Even though objective function (Equation (4.5)) requires increased computational efforts (more time and good initial guesses). That is the approach that has been applied in this work. The isoactivity criterion is used, when needed, in order to get good initial estimates for the mole fraction objective function. This approach has been previously proposed by several authors (Magnusson[72], Marcilla et al.[82]).

When needed, a regularization approach is used to accelerate and guarantee convergence. The expression for regularization remains the same and it is added to the expression in Equation (4.5):

$$OF = \sum_{i=1}^{I}\sum_{j=1}^{N}\left(x_{ij}^{cal} - x_{ij}^{exp}\right)^2 + w_{reg}\sum_{i=1}^{I}\sum_{j=1}^{N}\left(AIP_j - AIP_j^{IG}\right)^2 \tag{4.8}$$

where N is the number of experimental data points, AIP_j is the current value of the CI-interaction parameter j, AIP_j^{IG} its corresponding initial guess, and w_{reg} is a weighting value used to increase or decrease the influence of regularization in the optimization.

UNIFAC-CI (Dortmund) for reference model Modified UNIFAC (Dortmund)

For the development of this model, exclusively P-x_i-y_i and P-x_i VLE data were used. Whenever possible, VLE experimental data have been checked for thermodynamic consistency using the test of *Van Ness*[74]. In contrast to the UNIFAC-CI (VLE) model, an objective function using total pressures and activity coefficients (when needed) was used. One reason to use this approach is to get a fair comparison with the reference model[73], which uses the same objective functions for P-x_i-y_i and P-x_i data. It should be pointed out that UNIFAC-CI (Dortmund) model is not intended for prediction of γ^{∞}, H^E,

c_P^E, and LLE systems, since no experimental data related to these properties were used for regressing the AIPs.

The objective functions are defined as:

$$OF = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_{i-exp} - P_{i-calc}}{P_{i-exp}}\right)^2 \tag{4.9}$$

and/or

$$OF = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{P_{i-exp} - P_{i-calc}}{P_{i-exp}}\right)^2 + \left(\frac{\gamma_{i-exp} - \gamma_{i-calc}}{\gamma_{i-exp}}\right)^2\right] \tag{4.10}$$

where N is the number of experimental data points used for the parameter estimation. The required activity coefficients were calculated as follows:

$$\gamma_i = \frac{y_i P}{x_i P_i^{sat}} \tag{4.11}$$

where P is the total pressure in the system and P_i^{sat} is the pure component vapor pressure. The regularization proposed scheme is the following:

$$OF = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{P_{i-exp} - P_{i-calc}}{P_{i-exp}}\right)^2 + \left(\frac{\gamma_{i-exp} - \gamma_{i-calc}}{\gamma_{i-exp}}\right)^2\right] + w_{reg}\sum_{i=1}^{N}\left(AIP_i - AIP_i^{IG}\right)^2 \tag{4.12}$$

where N is the number of experimental data points, AIP_j is the current value of the CI-interaction parameter j, AIP_j^{IG} its corresponding initial guess, and w_{reg} is a weighting value used to increase or decrease the influence of regularization in the optimization.

4.3 EQUILIBRIUM DATA

Sufficient phase equilibrium data have been collected in order to regress the AIPs through experimental data. Since one of the goals of this work was to allow property predictions without the need for a large amount of experimental data (or without using new experimental data), moderate amounts of experimental data were selected for regressing the AIPs. A classification of the used phase equilibrium data is shown in Table 4.1.

Table 4.1 Type of Phase Equilibrium Data Used in the Correlation of Measured Data

Phase Equilibria Information	Type of data	Tests
VLE	P,x_i,y_i	Thermodynamic Consistency Test (Van Ness[74])
	P,x_i	None
LLE	T,x_i	None

4.3.1 Vapor-Liquid Equilibria

The VLE data consisted of more than 400 data sets. Most of the used data sets were extracted from the CAPEC database[83] and a small portion from the open literature. The time span covers the data from 1930-2008. The experimental set-ups in these data sets are diverse: vapor recirculation still, ebulliometer, equilibrium cell, modified static apparatus, to name a few. Table 4.2 lists the statistics for the data sets used for parameter optimization in this work.

Whenever possible, the data were tested for thermodynamic consistency following the work of Van Ness[74], except P-xi data which could not be checked for thermodynamic consistency. The data not satisfying the Van Ness test were removed from the final data system used for regressing the parameters. Approximately 75% of the data were tested for thermodynamic consistency.

Table 4.2 Statistics for the VLE Data System for Generating the AIPs

	Data sets	Data points	Main Groups involved
Functional Groups	227	3946	CH_2, C=C, ACH, $ACCH_2$, OH, ACOH, CH_2CO, CHO, CCOO, HCOO, CH_2O, COOH, c-CH_2, c-CH_2O, HCOOH
Methanol	36	636	CH_2, C=C, OH, H_2O, CH_2CO, CH_2O, COOH
Water	40	515	CH_2, OH, CH_3OH, CH_2CO, CHO, CCOO, CH_2O
N-atom related groups	59	855	CH_2, ACH, $ACCH_2$, OH, CH_2CO, CCOO, CNH_2, CNH, $(C)_3N$, $ACNH_2$, PYR, CCN, CNO_2

Cl-atom related groups	86	1325	CH_2, C=C, ACH, $ACCH_2$, OH, CH_2CO, CCOO, CH_2O, CCl, CCl_2, CCl_3, CCl_4, aCCl
S-atom related groups	19	434	CH_2, C=C, ACH, OH, CH_2CO, CCOO, CH_2O, CCl_3, CCl_4, DMSO, CH_2S

4.3.2 Liquid-Liquid Equilibria

The LLE data used consisted of about 37 binary data sets. The criteria for choosing LLE experimental data are somewhat arbitrary, there is not an equivalent and generally acceptable thermodynamic consistency test as used for VLE experimental data. At the same time, it is important to have as many data points as possible, covering the whole range of compositions. The election of binary data sets is based on the need for simplicity in the parameter optimization step in our work. Inclusion of ternary experimental LLE data has been planned for future work. The data sets have been collected from the DECHEMA experimental data collection[84]. The data used include compounds with C, O, and N atoms. Even though at present time there are binary LLE experimental data covering different systems of amines and aniline with water, those data sets have not been included in the data system used for correlation since groups CNH_2, CNH, $(C)_3N$, and $ACNH_2$ do not exist in the UNIFAC-LLE parameter table. In spite of this, with the regressed AIPs including C-N, O-N, N-N atom interactions (and mirror AIPs), interactions involving groups CNH_2, CNH, $(C)_3N$, and $ACNH_2$ can be potentially generated for evaluation. Table 4.3 depicts the statistics of the data sets used for parameter optimization in this work.

Table 4.3 Statistics of the LLE Data System for Generation of AIPs

	Data Sets	Data Points	Main Groups Involved
C,O-atom related groups	29	321	CH_2, OH, C=C, H_2O
N-atom related groups	8	62	CH_2, OH, H_2O, CCN, CNO_2, $ACON_2$

4.4 UNIFAC-CI (VLE) for reference model Original UNIFAC-VLE

4.4.1 Correlation Analysis

Regression Procedure

Before regressing the AIPs, it was needed to make a decision whether regression should be done in series or in parallel. Equivalences when regressing the GIPs in all the UNIFAC models are evident, since a dependency between the GIPs and AIPs clearly exists.

Historically, the GIPs in the UNIFAC models have been regressed in series. Fredenslund et al.[55] refrained from regressing all (or a significant amount of the total)

the GIPs in parallel mainly because the computational resources at that time were not powerful enough. Also, monitoring the optimization procedure was not considered practical (and this statement is still valid nowadays). In his work for UNIFAC-LLE, Magnusson[72] regressed the GIPs in parallel to avoid convergence problems and also 'insolvable' weighting problems due to presence of different kinds of mixtures and compounds in the experimental data sets. Following the same reasoning, Abildskov[1] regressed the KT-UNIFAC model GIPs using a parallel procedure in terms of group solutions going from 2 group solutions (for example, CH_2-C=C, CH_2-ACH, CH_2-$ACCH_2$ interactions) and beyond. However, Abildskov clearly states that the build-up procedure used in his work (parameters are regressed in a sequential manner where the results from previous parameter regressions are used as initial guesses for the additions of new parameters) can be considered as a mixed procedure '*keeping the advantages of fitting in series while fitting in parallel'*.

From the mathematical point of view, the complexity and nonlinearity of the UNIFAC expressions lead to a nonlinear regression problem which most of the time results in a multi-modal optimization problem. Kratch et al.[85] studied the optimization of UNIFAC GIPs using evolutionary algorithms when regressing the GIPs parameters in a simultaneous and sequential manner. It was confirmed that one of the most common problems when regressing in series is the fact that incompatibilities between the GIPs might occur. The reason is that might happen that no parameter set can be found in order to deliver satisfactory results, due the fact that previously fitted parameters remain constant.

On the other hand, regression of all the parameters simultaneously was found to lead to a more complex search space with significant computational requirements for the employed optimization algorithms. Their work, when regressing a set of four GIPs in Modified UNIFAC (Dortmund), did not lead to a definite answer about which approach is more convenient, perhaps (again) due to the complexity of the scenario for parameter optimization when regressing UNIFAC GIPs. When performing extensive parameter optimization for UNIFAC models, most (if not all) of the published works use local optimization algorithms (Nelder-Mead[86], Levenberg[87]-Marquadt[88]). No published work has been found using global optimization algorithms to regress GIPs of UNIFAC models.

In this work, a similar approach to that work used by of Abildskov[1] has been taken. As it was done for the GIPs in original UNIFAC, in this work, the AIPs were regressed first using data containing just C atoms (hydrogen, of course, is implicitly considered). In the same way, the UNIFAC matrix was built (from groups CH_2, C=C, ACH, $ACCH_2$…) the first AIPs regressed were b_{C-C}, c_{C-C}, d_{C-C}, d_{C-C}, $\overline{b_{C-C}}$, $\overline{c_{C-C}}$, $\overline{b_{C-C}}$, $\overline{b_{C-C}}$. This represents a good basis for further parameter optimization since interactions such as CH_2-C=C and CH_2-ACH are present in a great number of systems containing atoms other than carbon. Secondly, with all the C-C interactions fixed, the b_{C-O}, c_{C-O}, d_{C-O}, e_{C-O}, $\overline{b_{C-O}}$, $\overline{c_{C-O}}$, $\overline{d_{C-O}}$, $\overline{e_{C-O}}$ AIPs are regressed. Afterwards, the interactions O-O, C-N, O-N, N-N, C-Cl, O-Cl, N-Cl, Cl-Cl, C-S, O-S, N-S, Cl-S, and S-S were added. At the end (or in between the whole procedure if considered necessary) all the AIPs were regressed for simultaneous optimization. The optimization procedure is illustrated in Figure 4.1.

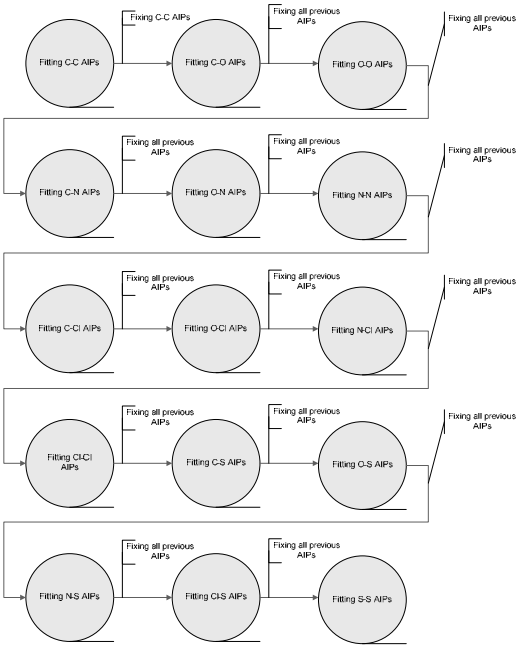


Figure 4.1 Procedure for regressing AIPs for UNIFAC-CI VLE

Improvements

When regressing a very large number of parameters, there are always issues that should be addressed. As mentioned in section 4.2.2, when needed, a scheme of regularization in the objective function has been used in order to control the variance of the AIPs, and also to improve the speed of the calculations. Common values of the regularization factor, w_{reg} (see Equation 4.3), lie in the range between $1x10^2$ and $1x10^8$. When good results could not be obtained, these regularization values were changed and the optimization was performed again. Besides performing an 'in parallel' optimization; inside of each subsystem for optimization, similar compounds were regressed step by step following the principle of 'solutions', used by *Abildskov*[1]. As an example, for C-O atom interactions, UNIFAC groups 1-5 (CH_2-OH) were regressed, followed by UNIFAC groups 1-8 (CH_2-ACOH), and UNIFAC groups 1-9 (CH2-CH_2CO). This was done in a progressive order in terms the UNIFAC main groups and also based on the number of n-solution systems (2 groups, 3 groups, etc).

On the other hand, it has been pointed out before that other empirical approaches[89] when defining the CI values of the main groups can lead to better results of the optimized AIPs. The reason seems to be that assymetric differences in the 'size' of the UNIFAC groups lead to high (either positive or negative) values of the GIPs. This happens because the difference in the CI values of the UNIFAC groups is also increased and the functionality (Equations (3.1) and (3.2)), that the GIPs have with the CIs values, is translated in disproportional GIPs values. These numbers can affect the quality of the correlation, and that is why modifying the way the connectivity indices are calculated was considered.

For instance, regressing the interactions involving the ACH main group, considering just the zeroth order CI, was not as good as expected. Therefore, as the group ACH is used to represent aromatic rings, its 0^{th} and 1^{st}-order connectivity indices were calculated assuming that each aromatic carbon is attached to two similar groups, and that these three atoms/vertices were included in the same group. This is not the case for the second-order connectivity index. This approach improved significantly the correlation results.

4.4.2 Comparison with the Reference Model

When correlating the experimental data, a measure of the error was chosen to be the average absolute relative deviation (AARD), defined in the following way:

$$AARD(\%) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_{\exp} - P_{calc}}{P_{\exp}}\right| \times 100 \qquad (4.13)$$

AARD (%) was chosen since it represents a fair way to take into account differences with respect to experimental data, regardless of the magnitudes in pressure units.

In the same way, as in original UNIFAC[55], two special tables of AIPs have been created for methanol and water (which are molecular groups), which means that for any calculation involving methanol or water the respective table of AIPs needs to be used to generate the GIPs. Correlation results are summarized in Figure 4.2 and Table 4.4, where there is a classification depending on the kind of compounds treated: functional compounds (containing C and O atoms), methanol, water, N-atom, Cl-atom and S-atom containing compounds.

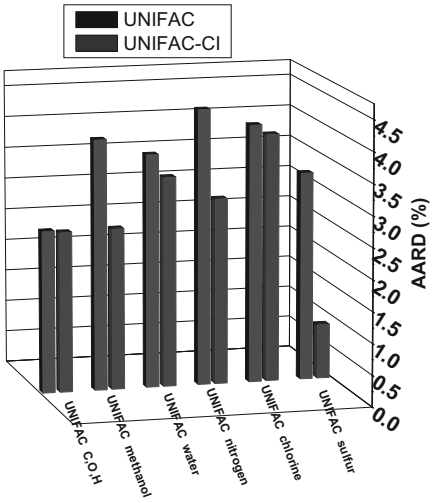


Figure 4.2 Correlation Statistics: Comparison of AARD with the Reference Model (Original UNIFAC VLE)

Table 4.4 Deviation Statistics between UNIFAC-CI and the reference model for all the subsystems considered in this work

	AARD% Original UNIFAC	AARD% UNIFAC-CI
C,O,H systems	2.57	2.54
Methanol	3.89	2.98
Water	3.40	3.38
N-atom systems	4.39	2.96
Cl-atom systems	4.11	3.85
S-atom systems	3.31	1.12

In Figure 4.3, results from Table 4.4 are represented through the individual deviations for each data set. The results are ordered in ascending order (in terms of AARD), with respect to UNIFAC-CI.

Surprisingly, for all the subsystems, the AARD is lower for the UNIFAC-CI model. The magnitude of the improvements depends of several factors, such as, the number of data sets used, if some of the data sets used for the AIPs parameter estimation were not used in original UNIFAC, and the kind of subsystems. It can be seen in Figure 4.3a clear that the C, O, H subsystem gave similar AARD. One reason for this is that the amount and diversity of data made it difficult to lower the deviations in comparison to the reference UNIFAC model. For example, the treatment of acids was particularly complicated for the C, O, H related systems. *However, it is important to mention that the objective of the UNIFAC-CI method is not to improve or substitute the reference UNIFAC models but to extend its application range with acceptable accuracy.*

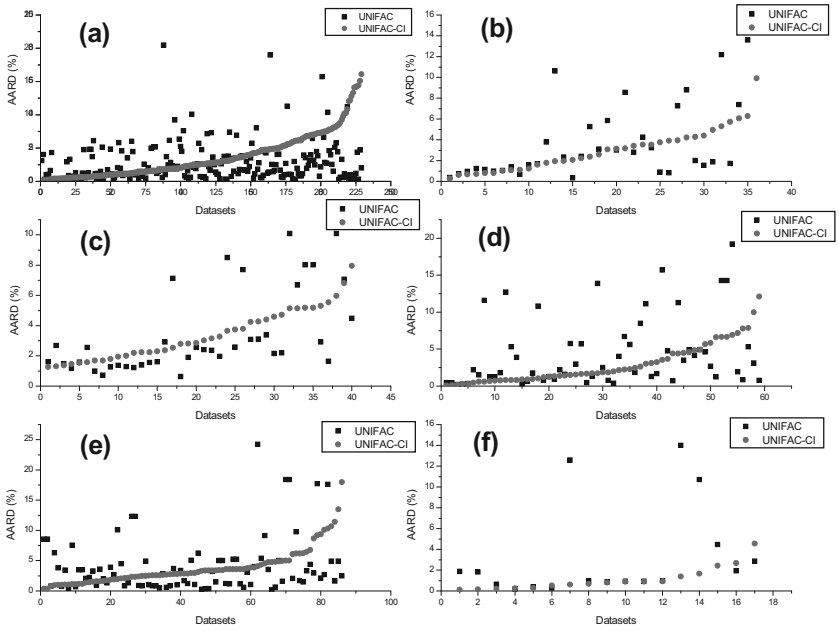


Figure 4.3 Comparisons of the AARD% values between UNIFAC-CI and original UNIFAC, for the subsystems: (a) C, O, H related sytems, (b) Methanol systems (C, O, H), (c) Aqueous systems (C, O, H), (d) N-atom systems, (e) Cl-atom systems and (d) S-atom systems.

4.5 ORIGINAL UNIFAC FOR LLE

4.5.1 Correlation Analysis

The parameter optimization for LLE systems was done using only binary system data and the main objectives were to generate the missing GIPs for the reference UNIFAC model parameter table proposed by Magnusson[72], and, to explore the goodness of the results from a qualitative point of view.

Regression Procedure

The parameter optimization was performed in the same way as it was done for UNIFAC-CI VLE. Since the interaction CH_2-H_2O is the most important because water is present in three quarters of the available LLE experimental data[84], and due to the fact that the CH_2 group is present in the majority of the published LLE data, systems including exclusively this interaction (i.e. alkanes-water) were considered first. This means that the following AIPs were regressed according to Equations (3.7) and (3.8): $b_{C-O}, c_{C-O}, d_{C-O}, e_{C-O}, \overline{b_{C-O}}, \overline{c_{C-O}}, \overline{d_{C-O}}, \overline{e_{C-O}}, b_{O-C}, c_{O-C}, d_{O-C}, e_{O-C}, \overline{b_{O-C}}, \overline{c_{O-C}}, \overline{d_{O-C}}, \overline{e_{O-C}}$. Once these AIPs were regressed, systems involving interactions with the groups CH_2, OH and H_2O (for example alcohol-water systems) were added to the previous data system. These systems were chosen because they are very abundant in the literature and the atom interaction O-O is considered very important for further calculations (systems involving alcohols and water were a good source to regress O-O AIPs). The following AIPs where therefore regressed: $b_{C-O}, c_{C-O}, d_{C-O}, e_{C-O}, \overline{b_{C-O}}, \overline{c_{C-O}}, \overline{d_{C-O}}, \overline{e_{C-O}}, b_{O-C}, c_{O-C}, d_{O-C}, e_{O-C}, \overline{b_{O-C}}, \overline{c_{O-C}}, \overline{d_{O-C}}, \overline{e_{O-C}}, b_{O-O}, c_{O-O}, d_{O-O}, e_{O-O}, \overline{b_{O-O}}, \overline{c_{O-O}}, \overline{d_{O-O}}, \overline{e_{O-O}}$. Later, systems involving nitrogen atoms (groups CCN, CNO_2, and $ACNO_2$) were introduced with the new group interactions H_2O-CCN, H_2O-CNO_2, H_2O-$ACNO_2$ present in the data system. In this case, the AIPs $b_{C-O}, c_{C-O}, d_{C-O}, e_{C-O}, \overline{b_{C-O}}, \overline{c_{C-O}}, \overline{d_{C-O}}, \overline{e_{C-O}}, b_{O-C}, c_{O-C}, d_{O-C}, e_{O-C}, \overline{b_{O-C}}, \overline{c_{O-C}}, \overline{d_{O-C}}, \overline{e_{O-C}}, b_{O-O}, c_{O-O}, d_{O-O}, e_{O-O}, \overline{b_{O-O}}, \overline{c_{O-O}}, \overline{d_{O-O}}, \overline{e_{O-O}}$ were fixed in order not to be perturbed (based in a larger database) when introducing new data sets; and the AIPs $b_{O-N}, c_{O-N}, d_{O-N}, e_{O-N}, \overline{b_{O-N}}, \overline{c_{O-N}}, \overline{d_{O-N}}, \overline{e_{O-N}}, b_{N-O}, c_{N-O}, d_{N-O}, e_{N-O}, \overline{b_{N-O}}, \overline{c_{N-O}}, \overline{d_{N-O}}, \overline{e_{N-O}}$ were regressed. At the end, all the AIPs were 'released' for parameter optimization. A scheme of the parameter optimization procedure is shown in Figure 4.4.

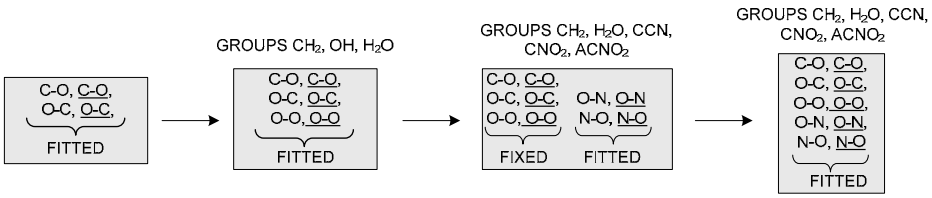


Figure 4.4 Schematic representation of the regressing procedure for the AIPs in UNIFAC-CI (LLE)

In the same way as Magnusson[72] did, when estimating the GIPs, it has been preferred to match as best as possible the mole fractions in a region of concentration where the two

liquid phases coexist, rather than to focus solely on precise predictions at small concentrations.

Limitations

As in Original UNIFAC (LLE), also in this work no temperature dependence of the interactions was considered. Therefore, safe extrapolation, with respect to temperature, cannot be guaranteed. Also the selection of the experimental data sets used for regressing the AIPs followed no specific criteria.

4.5.2 Comparison with the Reference Model

A qualitative agreement with experimental data has been the main objective for developing the UNIFAC-CI (LLE) model. For comparison purposes, and as a measure of 'goodness' of the parameter optimization, the UNIFAC-CI LLE model was compared against the reference model (Original UNIFAC-LLE) by inspection of the correlation results in the LLE phase diagrams. Each of the data sets were compared against the reference method, and whenever unsatisfactory results were obtained (e.g. high discrepancies), the weight factors were applied to those data sets with high deviations. It should be mentioned that no solution could have been found for some optimization scenarios. Problems, such as, very long convergence times, inadequate initial guesses, etc., were encountered. At times, during the optimization procedure, some data sets had to be excluded, for example, data sets with reocurring high deviations after several optimization runs.

Comparisons of correlation results for phase equilibria between hydrocarbons and alcohols with water, are shown in Figures 4.5 and 4.6. Data representing diluted concentration regions, as well as data representing the entire range of compositions have been chosen for these comparisons. Systems for which UNIFAC-CI (LLE) performs both 'better' and 'worse' than the reference UNIFAC (LLE) model have been selected, in order to make a critical discussion of the correlation results. For the systems pentane/water and 2-hexanol-water, significant differences between the CI model and the reference UNIFAC could not be observed.

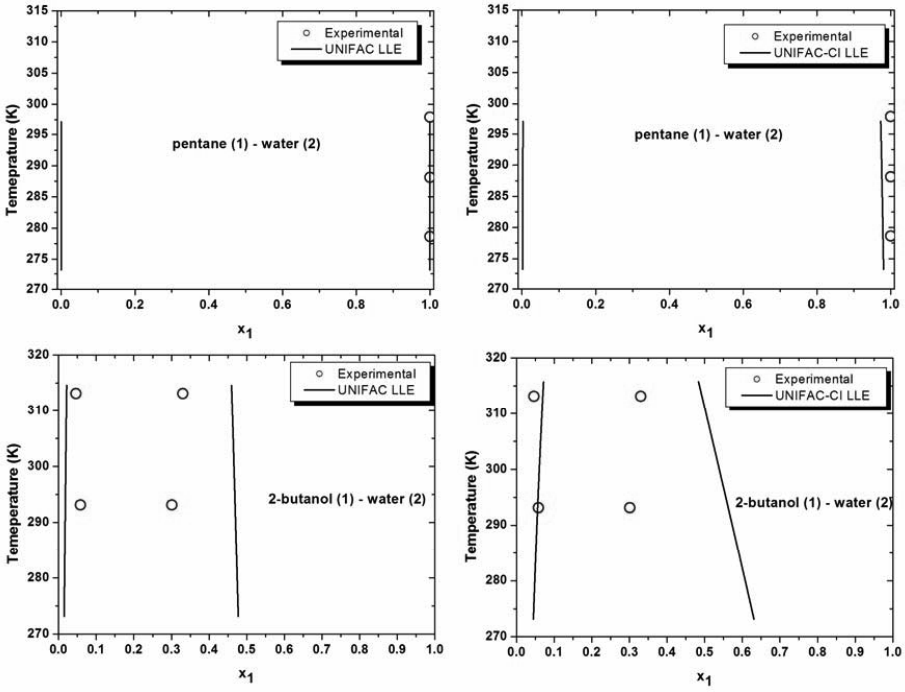


Figure 4.5 LLE Phase Diagrams. Comparisons of correlation results for Original UNIFAC (LLE) and Original UNIFAC-CI (LLE) for the systems: pentane/water and 2-butanol/water.

On the other hand, for the systems 1-butanol/water and 2-butanol/water, the discrepancies are more pronounced. It should be noted that for Original UNIFAC (LLE) model, the generated phase diagrams (for 1-butanol/water and 2-butanol/water) follow almost exactly the same trend because the same group interactions and GIPs were used in both systems. Therefore, the 'prediction' is the same even though the experimental phase equilibria is different for water with 1-butanol or 2-butanol. UNIFAC-CI (LLE) does the same. This is one case that shows that UNIFAC-CI cannot improve predictions for which the reference model (Original UNIFAC (LLE)) shows deficiencies, and that is not the objective of this work.

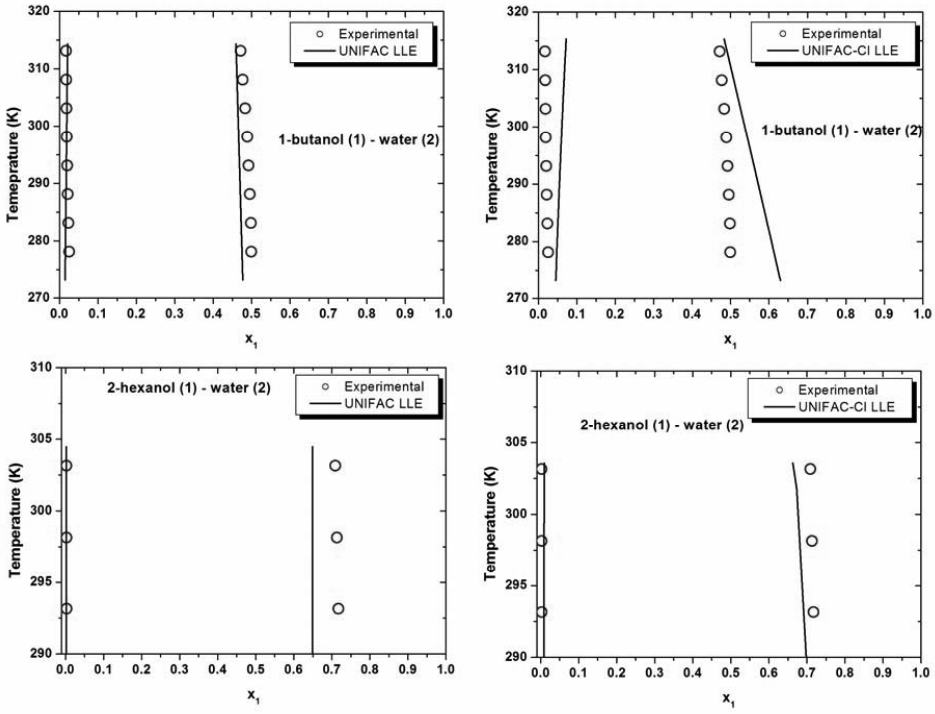


Figure 4.6 LLE Phase Diagrams .Comparisons of correlation results for Original UNIFAC (LLE) and Original UNIFAC-CI (LLE) for the systems: 1-butanol/water, and 2-hexanol/water.

When discussing how 'good' the correlations are when comparing both models, it can be said that the reference model (Original UNIFAC (LLE)) is doing slightly better for the correlation of the experimental data shown in Figures 4.5 and 4.6, however for the purposes of this work the obtained results are considered as a good starting point for future investigations using the UNIFAC-CI (LLE) model.

Figures 4.7 and 4.8 highlight the comparisons of LLE phase diagrams for systems involving nitrogen-containing compounds and water. For these cases, differences between experimental and calculated values were observed to be similar to those using original UNIFAC (LLE). For the systems propionitrile/water and adiponitrile/water (involving the interaction CCN-H_2O), where the experimental data over the whole range of concentration are available, it should be noted that the parameters of UNIFAC-CI LLE also 'fail' to represent the phase splitting, just as the reference model. For the system nitrobenzene/water (involving the $ACNO_2$-H_2O UNIFAC group), the CI-model gives better results at high nitrobenzene concentration.

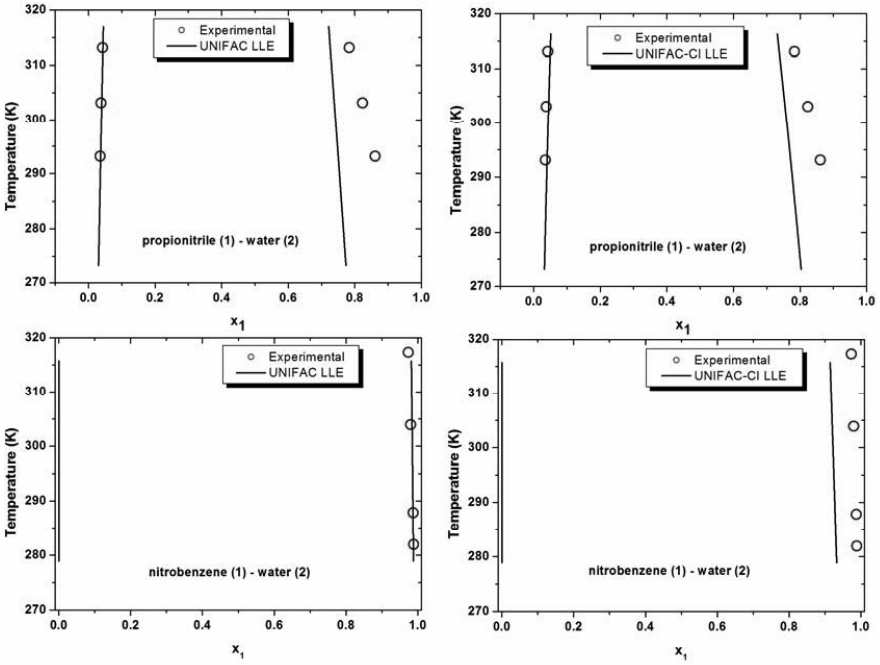


Figure 4.7 LLE Phase Diagrams. Comparisons of correlation results for Original UNIFAC (LLE) and Original UNIFAC-CI (LLE) for the systems: propionitrile/water and nitrobenzene/water.

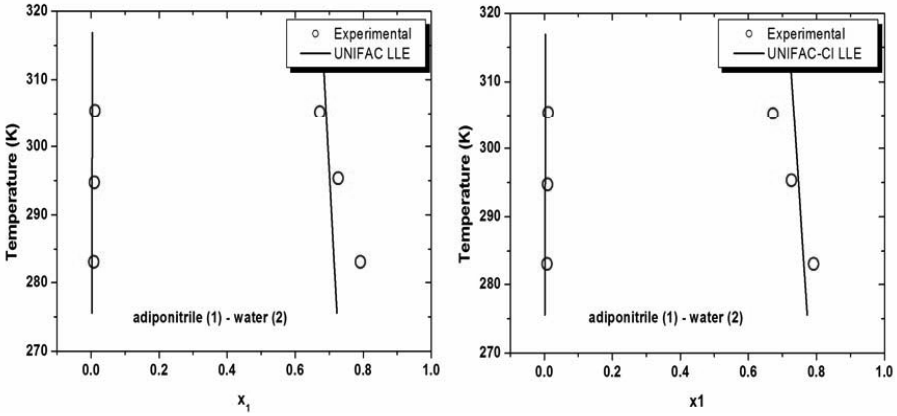


Figure 4.8 LLE Phase Diagrams. Comparisons of correlation results for Original UNIFAC (LLE) and Original UNIFAC-CI (LLE) for the system: adiponitirile/water.

In general, when comparing the LLE phase diagrams (for systems used in the parameter optimization) to those obtained for the reference model, results similar to those shown

in Figures (4.5)-(4.8) can be expected. So far, the results show that an split success has been achieved. More work is necessary in order to fine-tune the AIPs: work related to numerical aspects of the parameter optimization procedure, model calibration-reformulation among other topics. The correlation statistics can be found in Appendix D for the individual data sets used in the correlations of the AIPs.

4.6 UNIFAC-CI (DORTMUND) FOR VLE

4.6.1 Correlation Analysis

The data system of experimental VLE data used for the generation of the AIPs for Modified UNIFAC-CI (Dortmund) is the same as the one used for regression of the parameters for Original UNIFAC-CI, except for chlorine and sulfur related compounds which were not considered in this part of the work. As stated in section 4.1, the AIPs were regressed through the experimental data available in the database (and also used for UNIFAC-CI (VLE) model – see Table 4.2).

Regression Procedure

The regression procedure was performed in the same manner as for the original UNIFAC-CI (VLE). The procedure outlined in Figure 4.1, with the exception of groups with chlorine and sulfur atoms, has been used.

Improvements

When necessary, a regularization scheme has been applied with regularization factor values ranging between $1x10^2$ and $1x10^4$. Systems involving alcohols, furans, acids, water and methanol were particularly difficult to use in the regression as they slowed down the optimization calculations, thereby making it necessary to use high levels of regularization to guarantee convergence.

The 'solutions' approach was applied when introducing the ordered data sets for regression. Unlike the regression for the Original UNIFAC-CI (VLE), the following groups: (42) c-CH2, (43) c-CH2O, and (44) HCOOH, were included. Also, empirical descriptions of CI-description for the Modified UNIFAC-CI (Dortmund) groups were explored with the objective of improving the correlation results. In Figure 4.9, the new description of CIs for the main group ACOH is presented.

Figure 4.9 Connectivity Indices Representation for the UNIFAC main Group ACOH. Description used for Original UNIFAC-CI (left) and Modified UNIFAC-CI (Dortmund) (right).

Arbitrarily, In Original UNIFAC-CI (VLE) both atoms (carbon and oxygen) have the delta values corresponding to acyclic atoms (following the description of Kier-Hall[59]). For modified UNIFAC-CI (Dortmund) the carbon atom has a characteristic delta value for cyclic carbons (14-nH), which would be the natural description in terms of CIs. These CI descriptions gave better correlation of the data for each of the UNIFAC-CI models, and diminished the convergence times during the parameter optimization calculations. In general, there are no specific rules to follow when choosing the CIs description, but unless a lot of improvement can be achieved with empirically modified CIs values, caution should be paid for not breaking any basic rule of Kier-Hall´s theory when modifying CI descriptions.

4.6.2 Comparison with the Reference Model

For the correlation results, deviations with respect to experimental data were analyzed in terms of average absolute relative deviation (AARD), which has been defined in Equation (4.13). A parameter table for groups involving C, O, H, and N atoms; and a parameter table for the molecular group H_2O have been generated. Correlation results are summarized as deviations (in terms of AARD %) for each of the systems (functional and aqueous) used for regression, is shown in in Figure 4.10 and Table 4.5, where a comparison in terms of AARD is made between the reference model and the corresponding UNIFAC-CI model

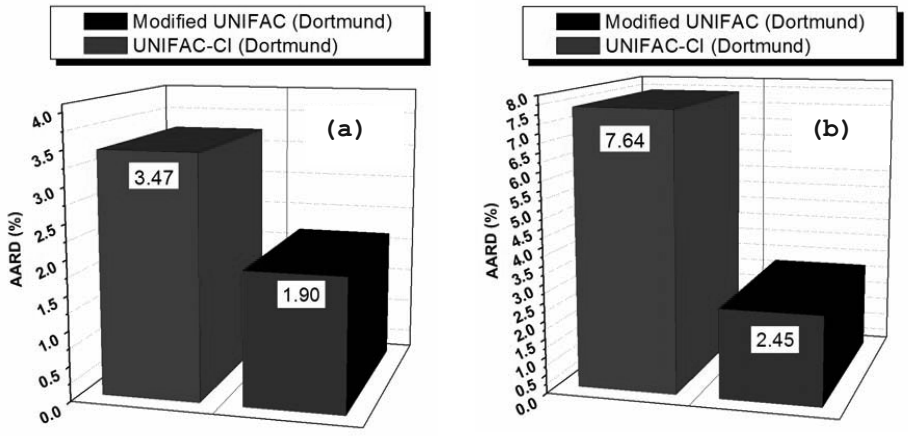


Figure 4.10 Graphical comparison of the overall AARD% between the UNIFAC-CI (Dortmund) model and the reference model (Modified UNIFAC Dortmund) for: a) the C, O, H and N systems used in the correlation, and b) aqueous systems used in the correlation.

Table 4.5 Correlation Statistics: Comparison of AARD with the Reference Model (Modified UNIFAC Dortmund) for the C, O, H and N systems used in the correlation.

	$AARD_{\%}$ Ref. Model (Modified UNIFAC-Dortmund)	$AARD_{\%}$ CI-model
C, O, H, N systems	1.90	3.47
Water	2.45	7.64

It can be seen in Table 4.5, that the performance of the regressed UNIFAC-CI model for the reference Modified UNIFAC (Dortmund) model is acceptable for the C, O, H, N systems; but not satisfactory for the aqueous systems.

In Figures 4.11 and 4.12, deviations (in terms of AARD %) for each of the data sets used, are shown for both C, O, H, N related systems and aqueous systems. The deviations (AARD %) are sorted in the increasing order with respect to UNIFAC-CI (Dortmund). The highest deviations for the UNIFAC-CI Dortmund model (in Figure 4.11) belong to systems involving ethers, amines, acids, aldehydes and alcohols. However, not all the compounds belonging to these chemical families give high deviations. There is no evident reason as to why such high deviations for some of these compounds occur, specially because these systems did not exhibit high deviations for the reference model. An analysis to identify the causes for these deviations is not a straightforward task, since the modifications intended to improve the correlation (i.e., different regularization levels, CIs description, different initial guesses, etc.) do not consistently provide the right result.

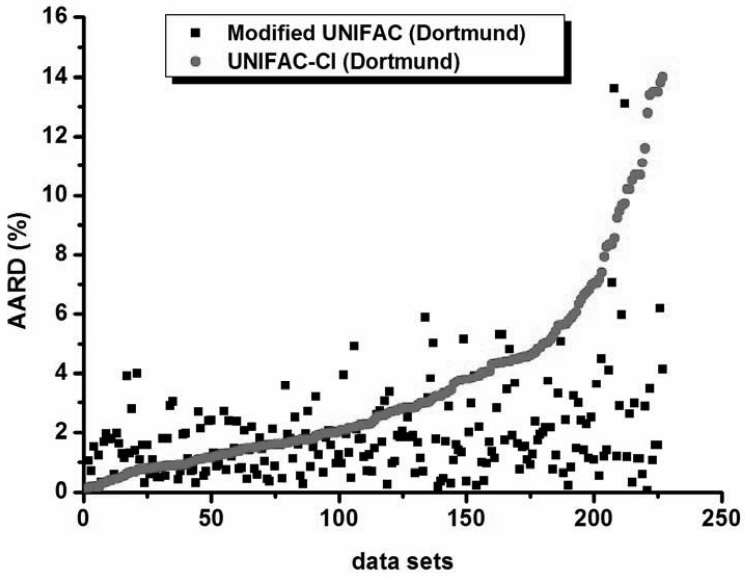


Figure 4.11 Comparisons of the AARD% values between Modified UNIFAC (Dortmund) and the CI model for systems involving C, O, H, and N atoms

The correlation results for the aqueous systems are shown in Figure 4.12. Most of the data sets correspond to systems involving alcohols and water. The correlation with the CI models is not satisfactory, with deviations for some systems exceeding 10%. An inspection of the data system used for correlation reveals that all the data sets, with deviations higher than 10%, are data sets involving the system ethanol/water. Several optimization schemes were used in order to decrease the deviation, however, no significant improvements were achieved. All other systems besides the ethanol-water system were found to be satisfactory from a semi-quantitative point of view, however,

considerations of how to improve the correlation for aqueous systems should be considered for future work.

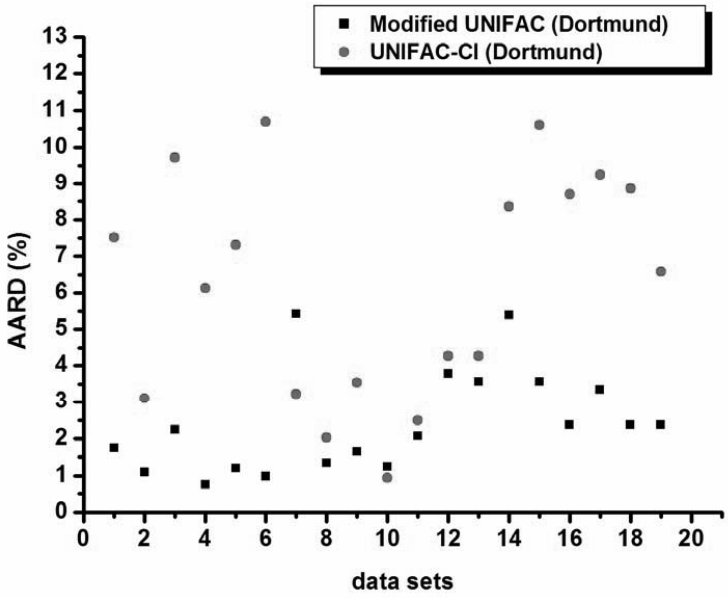


Figure 4.12 Comparisons of the AARD% values between Modified UNIFAC (Dortmund) and the CI model for Aqueous Systems

4.7 INADEQUACIES AND LIMITATIONS

The improvement in the performance of the UNIFAC models is not one of the priorities of this work, therefore, it is not reasonable to expect much better correlation results using the CI-models. The UNIFAC-CI models have not been developed to correct inadequacies in the reference model. Models such as Modified UNIFAC (Lyngby)[19], Modified UNIFAC (Dortmund)[73], and KT-UNIFAC[23] have been developed for such purposes.

Part of the strength of the UNIFAC models is that the large phase equilibria database used for regressing the GIPs, and adequate and careful parameterization of the UNIFAC models over the years (more than 30 years), have made the UNIFAC models an important tool for the chemical engineering community. In this work, it is believed that an extension in the application range (by generating the missing GIPs) does not necessarily mean that a large database is needed for regression. Before creating a bigger database, however, it is necessary to set up the rules and to identify potential problems and failures of the methodology. This is what has been done in this work.

What can be considered a limitation regarding the database used for regressing the AIPs is the homogeneity of the classes of compounds present in the data set. It has been

highlighted in Chapter 3 that one of the strengths of the CI-model is that, once the atomic parameters (for example, C-C, C-N, N-Cl, etc.) are regressed, any GIP including such atom interactions can potentially be generated. However, the smaller the number of UNIFAC groups considered when regressing a x-y atomic interaction (for example C-N) the less diversity and 'richness' in the information contained by such C-N atomic interaction when it comes to prediction of GIPs (see Chapter 5). For instance, in the estimation of nitrogen related AIPs for Original UNIFAC-CI (VLE), data sets containing the UNIFAC groups (27) $ACNO_2$, (36) ACRY, (39) DMF, (44) NMP, (46) CON, and (49) Morpholine were not taken into account when regressing the AIPs. This fact may influence the prediction capabilities of the UNIFAC-CI models for systems with compounds involving these UNIFAC group. Note that similar observations would also be made if the extrapolation of the group parameters were to be investigated.

4.8 CONCLUSIONS

According to the objectives of the GC^{Plus} methodology for filling out the UNIFAC parameter table, the group interaction parameters for different UNIFAC models generated through CIs have shown to give good correlation results for VLE and LLE. Due to the complexity and the amount of parameters needed to be regressed, different optimization schemes have been used in order to obtain the best possible parameters. The atom interaction parameters (that are necessary to calculate the group interaction parameters) were regressed by matching experimental VLE and LLE data.

Several investigations and optimization runs that are not reported here were performed 'backstage' to make sure that an optimal set of parameters could be obtained. A systematic approach was developed in order to regress the atom interaction parameters with a minimum consumption of computer time and with the highest possible reliability. Hence, the set of 'trained' atom interaction parameters is considered to be good enough for the generation of missing UNIFAC group interaction parameters for the original UNIFAC (VLE and LLE) and the Modified UNIFAC (Dortmund) models, and for the prediction of phase equilibria based on the UNIFAC-CI methodology.

Chapter 5

Predictions and Case Studies

5.1 INTRODUCTION

In Chapter 4, the correlation of data involving the UNIFAC-CI (VLE), UNIFAC-CI (LLE) and UNIFAC-CI (Dortmund) models was described together with the generation of AIPs needed to back-calculate some of the missing GIPs in the UNIFAC parameter tables. The compounds that were used for correlation and are expected to be described by these UNIFAC-CI models are the following:

Original UNIFAC-CI (VLE)

- Hydrocarbons
- Alcohols
- Water
- Ketones and aldehydes
- Ethers
- Esters
- Carboxylic acids
- Amines, Nitriles
- Chlorinated Compounds
- Sulfur-containing Compounds

Original UNIFAC-CI (LLE)

- Hydrocarbons
- Alcohols
- Water
- Amines and Nitriles

UNIFAC-CI (Dortmund)

- Hydrocarbons
- Alcohols
- Water
- Ketones and aldehydes
- Ethers
- Esters
- Carboxylic acids
- Amines, Nitriles.
-

Extrapolations to other kinds of systems (those involving only the AIPs corresponding to atoms studied in this work) are possible. However, even the GIPs for UNIFAC groups not used for parameter optimization can be generated if the needed AIPs are available. As an example, the group interactions involving the main group 46 (CON, amide) can be generated using the estimated AIPs for C, O and N atoms, and then used for property predictions. This can be attractive for property prediction and extrapolation

of the GIPs generated through CIs, however, such a demanding extrapolation is not recommended; results obtained in such a manner require careful interpretation and should not be taken for granted. Due to the high number of systems and combinations that can potentially be created using the CI-based GIPs for UNIFAC, it is difficult to assess for which systems the methodology will or will not give satisfactory predictions. Such an extensive analysis for assessing the boundaries of the UNIFAC-CI models in terms of the quality of predictions is planned for future work. It should be pointed out that for UNIFAC group interactions that were estimated based on only one or two data sets from the data system, there is an uncertainty about how well they would represent other systems. This uncertainty exists for the GC-model, and therefore, also for the UNIFAC-CI models.

A thorough statistical analysis of predictions and applications of the UNIFAC-CI and UNIFAC-CI (Dortmund) models falls outside the scope of this work. The purpose of this chapter is: 1) to show, for a limited but representative number of cases, satisfactory prediction results that can be expected using the present UNIFAC parameter matrices with the generated AIPs; 2) to illustrate, through three case studies, the possible applications, capabilities and restrictions of the UNIFAC-CI models; and 3) to discuss objectively the potential, limitations, and opportunities when using the UNIFAC-CI models.

The prediction results in this chapter are based exclusively on binary data that were not used for the parameter estimation procedure. As it was stated in Chapter 4, the VLE data used for correlation were based on isothermal data, but, for the predictions, both isobaric and isothermal data have been used. For both Original UNIFAC-CI and UNIFAC-CI (Dortmund) models, a comparison between experimental and calculated data is discussed in this chapter. Instead of showing the representation of all the systems predicted, some of the representative results will be shown. The tables detailing the statistics are given in the quantitative analysis sections, while several examples and specific comments on the results are placed in the qualitative analysis sections.

The case studies cover the extrapolations using the AIPs for predictions within the UNIFAC parameter table, as well as the evaluation of their use for solid-liquid equilibria (SLE) calculations. Finally, two examples where the UNIFAC-CI model is used to generate missing UNIFAC GIPs are also shown, together with a discussion.

5.2 PREDICTIONS OF MEASURED DATA

5.2.1 Original UNIFAC-CI

Quantitative Analysis

The criteria for comparison of the deviation of the UNIFAC-CI model with respect to the experimental data are defined as follows:

AARD % (Average Absolute Relative Deviation for pressure):

$$AARD(\%) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_{\exp} - P_{calc}}{P_{\exp}}\right| \times 100 \qquad (5.1)$$

SARE (Sum of Absolute Relative Error for pressure (P)):

$$SARE = \sum_{i=1}^{N} \left| \frac{P_{\exp} - P_{calc}}{P_{\exp}} \right| \tag{5.2}$$

and SRE (Sum of Relative Error for pressure (P)):

$$SRE = \sum_{i=1}^{N} \left(\frac{P_{\exp} - P_{calc}}{P_{\exp}} \right) \tag{5.3}$$

The relationship between SARE and SRE gives information on how systematic the deviations are. Whenever all the individual data points are underpredicted, SARE equals SRE and the same happens when all the data points are overpredicted.

In general, it can be stated that, the broader the data system used for a parameter optimization, the wider the safe range of the predictions that can be achieved. In Table 5.1, the statistics obtained for predictions of different binary systems that were not used for the parameter estimation can be found. Otherwise, all the predictions shown in this chapter are made using only GIPs generated through CIs, and not the original GIPs.

Table 5.1 Statistics for predictions using UNIFAC-CI. ND is the number of data points used per set.

System	ND	SARE	AARD	SRE
(1) trans-2-butene(2) 1-propanol (364K)	26	2.60	10.83	-2.60
(1) trans-2-butene (2) 2-butanol(364K)	27	1.68	6.73	-1.68
(1) trans-2-butene(2) 2-Methyl_2-propanol (364K)	27	0.85	3.41	-0.85
(1) trans-2-butene (2) 2-propanol (364K)	27	1.95	7.79	-1.95
(1) butylamine (2) hexane (60C)	7	0.55	11.07	0.55
(1) DMSO (2) acetone (25C)	11	0.06	0.69	-0.06
(1) n-heptane (2) monochlorobenzene (323K)	18	4.25	26.59	-4.25
(1) n-heptane (2) monochlorobenzene (343K)	12	1.97	19.66	-1.97
(1) ethyl-acetate (2) monochlorobenzene (313K)	15	0.85	6.53	0.85
(1) ethyl-acetate (2) monochlorobenzene (353K)	15	0.84	6.43	0.84
(1) ethyl-acetate (2) monochlorobenzene (393K)	15	0.94	7.22	0.94
(1) 1-pentene (2) monochlorobenzene (280K)	14	3.22	26.85	-3.22
(1) n-butyl-chloride (2) benzene (348K)	11	1.13	12.59	1.13
(1) n-butyl-chloride (2) benzene (398K)	11	1.08	12.03	1.08
(1) acetonitrile (2) 1-propanol (66.6kPa)	16	0.64	4.60	0.64
(1) acetonitrile (2) 1-propanol (101.3kPa)	16	0.46	3.26	0.46
(1) acetonitrile (2) 1-propanol (53.32kPa)	16	0.82	5.82	0.82
(1) acetonitrile (2) 1-propanol (39.99kPa)	16	1.04	7.42	1.04
(1) CCl4 (2) hexane (760mmHg)	18	0.72	4.49	0.72
(1) CCl4 (2) octene (760mmHg)	18	0.59	3.71	0.59
(1) n-hexane (2) 1-decene (101.3kPa)	21	0.30	1.60	0.30
(1) n-octane (2) 1-decene (101.3kPa)	36	0.07	0.22	0.07
(1) hexane (2) aniline (101kPa)	15	3.98	30.63	3.98
(1) 1,2-dichloroethane (2)_benzene_(298K)	13	3.00	27.31	3.00

As it can be seen in Table 5.1, in most of the cases the pressure values are well correlated, except for several exceptions (even though is difficult to establish conclusions from a limited data system for predictions) - such as some of the amine/alkane, alkene/alcohol, and monochlorobenzene/hydrocarbon systems.

The predictions of the UNIFAC-CI (VLE) model are reliable for systems that include compounds that have been used previously for parameter optimization. This means that

for such kind of systems, significant quantitative discrepancies with experimental data are not expected to be found. At this stage of the project, it is not possible to assess the capabilities of the model for very different systems from those the CI-model has been tested for correlation and prediction mode.

Qualitative Analysis

In all of the predicted systems using UNIFAC-CI (VLE), the phase diagrams were inspected in order to evaluate systematic trends in deviations with respect to measured data. Some representative examples are shown below in order to give a flavor of the capabilities and limitations of the CI-model.

Systems involving C, O, and H atoms

Isothermal predictions of alkene-alcohol systems when using GIPs generated from CIs, can be seen in Figures 5.1a and 5.1b. The agreement with experimental data is very good, and the difference in the quality of the prediction between the systems is small. Close observation, however, reveals slightly lower accuracy for the 2-butanol system.

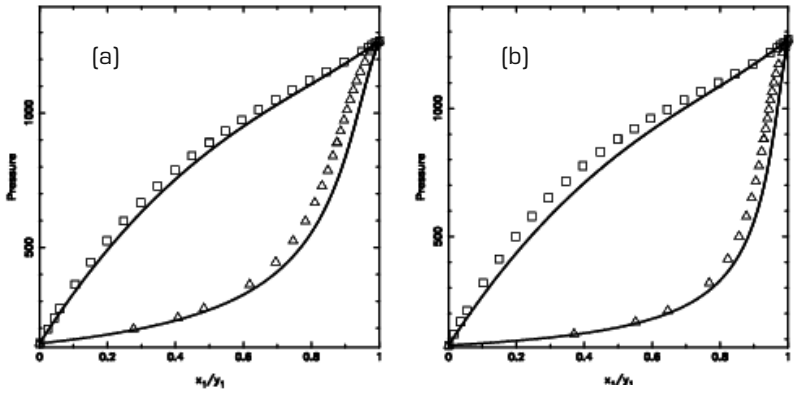


*Figure 5.1 UNIFAC-CI VLE prediction for the systems: (a) trans-2 -butene (1) / 2-methyl-2-propanol (2) and (b) trans-2-butene (1) / 2-butanol (2). T = 364.15 K. P=kPa. - Δ (measured). Data taken from: Roininen, J. Chem. Eng. Data., **53**, 207 (2008).*

An example of an isobaric prediction can be seen in Figure 5.2, for the system n-octane/1-decene at 760 mmHg. The agreement was, as expected, found to be very good, due to a large number of data available for the optimization procedure for the UNIFAC group interactions involved (CH_2-C=C, CH_2-OH. C=C-OH, etc.).

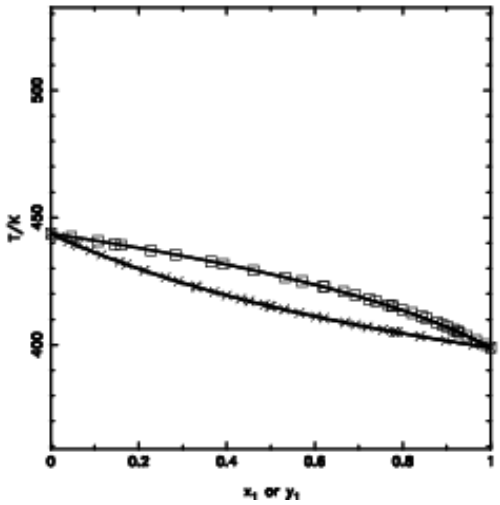


Figure 5.2 UNIFAC-CI VLE prediction for the system n-octane (1) / 1-decene (2). P = 760 mmHg. -× (measured). Data taken from: Hossein, J. Chem. Eng. Data, ***398****, 402 (2008).*

Systems with nitrogen-containing compounds

Systems involving the interactions CH_2-$ACNH_2$, CH_2-CCN, and OH-CCN are represented in Figure 5.3. Both Figures 5.3a and 5.3b are isobaric phase diagrams. In Figure 5.3a there is a disagreement between the prediction and the experimental temperatures at low concentrations of n-hexane. The reason for this could be the type of the aniline systems used in the parameter optimization: aniline-alkanes were not used, and probably the system n-hexane/aniline is greatly influenced by the regressed AIPs using the CH_2-$ACNH_2$ interaction present in other systems. On the other hand, the T-x phase diagram for the system acetonitrile/1-propanol at 66.6 kPa, shown in Figure 5.3b, shows good agreement between the experimental data and the CI model. It should be pointed out that the systems involving nitrogen-containing compound-alcohol were particularly difficult systems to correlate and, in general, the predictions of these systems are expected to be moderately good from the qualitative point of view. Acetonitrile/1-propanol could be considered a good achievement and a pleasant surprise.

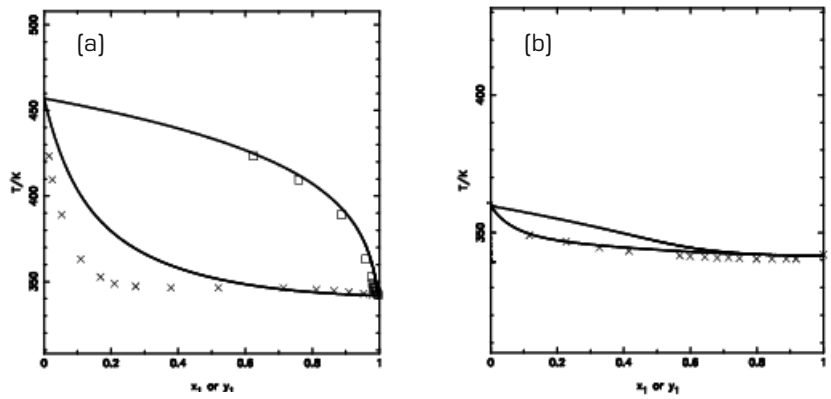


*Figure 5.3 UNIFAC-CI VLE prediction for the systems (a) n-hexane (1) / aniline (2) . P = 760 mmHg. -× (measured). Data taken from Gupta S.K., Nanotl S.M., and Rawat B.S., J. Chem. Eng. Data, **37**, 163 (1992). (b) acetonitrile (1) / 1-propanol (2). P=66.65 kPa. × (measured). Data taken from: Prasad et al. Can. J. Chem. Eng., **64**, 813 (1986)*

Systems with chlorinated compounds

Examples of phase diagrams for systems with chlorinated compounds predicted using GIPs derived from CIs are given in Figures 5.4a and 5.4b. For the system methyl-acetate/monochlorobenzene at 353.15 K, the agreement with the experimental pressures is very good, but, the near ideal behaviour of the system is not very difficult to predict.

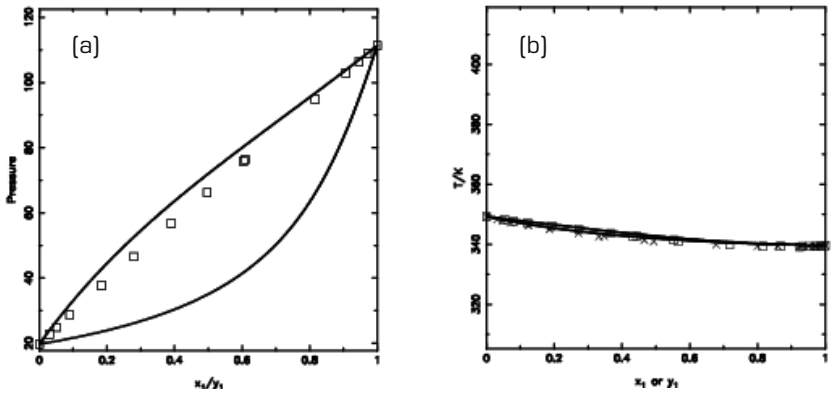


*Figure 5.4 UNIFAC-CI VLE predictions for the systems: (a) ethyl-acetate (1) / monochlorobenzene (2). T = 353 K. P=kPa. □ (measured). Data taken from: Maher P.J. and Smith B.D., J. Chem. Eng. Data, **24**, 363 (1979); (b) n-hexane (1) / n-butyl-chloride (2). P = 94.4 kPa. □ -× (measured). Data taken from: Wisniak J., and Kunis A., JCED, **40**, 1146*

For the system n-hexane/n-butyl-chloride at 94.4 kPa, the agreement with the experimental data is good. In general, however, it is necessary to be cautious when relying on the results of predictions for other mixtures involving chlorine-containing compounds, given that a small amount of experimental data was used for parameter estimation.

Systems with sulfur-containing compounds

Out of all the experimental data gathered for systems with sulfur-containing compounds, a small portion was preserved only for the purpose of validation of the predictions using UNIFAC-CI (VLE), and the other part was used for correlation of the S-atom containing AIPs. In general, as it was already pointed out when analyzing predictions of systems with nitrogen-containing compounds, systems including alcohols are complicated to correlate and to predict. This is shown in Figure 5.5a, where a comparison between the experimental data for 2-methyl-1-propanol/dimethyl-sulfide system and the prediction using UNIFAC-CI is given.

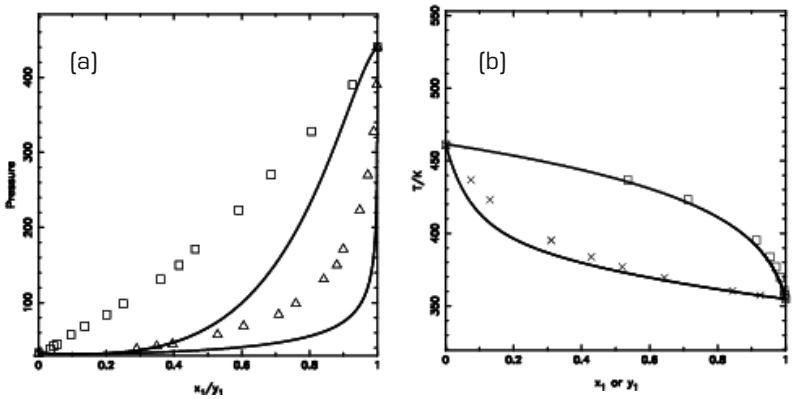


Figure 5.5 UNIFAC-CI VLE predictions for the systems: (a) 2-methyl-1-propanol (1) – dimethyl-sulfide (2). T = 365.65 K. P= mmHg. □-Δ (measured). Data taken from: Betancourt T. and McMillan A.F., J. Chem. Eng. Data, ***17****, 311 (1972); (b) 1,2-dichloroethane (1) – DMSO (2). P = 714.75 mmHg. □ -× (measured). Data taken from: Radhamma et al., J. Chem. Eng. Data,* ***53****, 374 (2008).*

It can be seen in Figure 5.5a that the experimental pressure is under-predicted, posing a problem -specially- in the low alcohol concentration region, most likely predicting an unexpected azeotrope. An explanation for such a discrepancy between measured and predicted data could be that the required OH-CH_2S interaction has not been used for regressing the AIPs.

The prediction in Figure 5.5b is a special case of using a 'parallel' AIPs parameter table (Appendix B, Table B.7) for the DMSO-related group interactions (i.e. CCl-DMSO). In this case, for 1, 2-dichloroethane/DMSO, the prediction is good. This system will also be considered in one of the case studies given below.

Conclusions

One of the most attractive features of the UNIFAC-CI models is that with the CI-generated AIPs, potentially any UNIFAC GIP (so far for those groups involving C, O, N, Cl and S atoms) involving such interactions can be calculated. However, it is clear that extrapolations 'far' outside the region of correlation (i.e., the UNIFAC groups used for the parameter optimization) have a high degree of uncertainty. The quality of the interpolation of predictions (for example, the one described in Figure 5.5a) is yet to be analyzed. A parameter optimization scheme addressing the observed limitations, together with the use of a larger and selected amount of new experimental data covering the key GIPs, necessary to guarantee safe predictions in the required application range is needed. It is important to identify trends for the 'good' and 'bad' results and the compound classes for which they are observed, in order to create safe guidelines for the user and, in order to refine and improve the methodology – work on this is still in progress at the Computer-Aided Process/Product Engineering Centre.

The temperature range for the VLE experimental data used for obtaining the AIPs was 273-400 K, and the pressures were below 10-15 atm. It is evident that the same recommendation, as given by the UNIFAC developers regarding temperatures and pressures for predictions, should be pointed out: extrapolations outside the temperature and pressure ranges of the experimental data systems used for regression, can lead to uncertain results. Not much can be said about the range of temperatures and pressures for predictions where the GIPs generated through CI can be applied. However, even though the valence connectivity indices do not have temperature functionality, and that missing (and new) GIPs can be generated through the GC^{Plus} methodology, it does not necessarily mean that the same recommended ranges of temperature and pressure should not be respected.

5.2.2 UNIFAC-CI (Dortmund)

The correlation statistics shown in Chapter 4 (section 4.6) indicate the need for a further refinement of the correlation for the data system used to regress the AIPs. Nevertheless, a limited number of systems have been checked for predictions using the AIPs corresponding to the UNIFAC-CI (Dortmund) model. Extreme extrapolations of the predictions have, in general, been avoided, but, some examples of extrapolation are shown in the qualitative analysis section.

Quantitative Analysis

The deviation criteria are the same as defined for UNIFAC-CI model (i.e., SAR, SARE, and AARD) in section 5.2.1. The comparison between several experimental data sets and the predictions using UNIFAC-CI (Dortmund) is given in Table 5.2

Table 5.2 Statistics for Predictions Using UNIFAC-CI (Dortmund). ND stands for number of data points in the data set.

System	ND	SARE	AARD	SRE
(1) acetic-acid (2) n-octane (323K)	21	0.86	4.51	-0.79
(1) acetic-acid (2) n-octane (343K)	20	0.47	2.61	-0.24
(1) trans-2-butene (2) 1-propanol (364K)	26	3.88	16.16	-3.88
(1) trans-2-butene (2) 2-butanol (364K)	27	3.77	15.10	-3.77
(1) trans-2-butene (2) 2-propanol (364K)	27	4.00	15.99	-4.00
(1) trans-2-butene (2) 2-methyl-2-propanol (364K)	27	3.80	15.20	-3.80
(1) 1-propanol (2) di-n-propyl-ether (278K)	15	0.61	4.70	0.61
(1) 1-propanol (2) di-n-propyl-ether (318K)	15	0.36	2.79	0.36
(1) 1-propanol (2) di-n-propyl-ether (323K)	15	0.23	1.78	0.21
(1) isopropanol (2) di-n-propyl-ether (288K)	14	0.28	2.32	-0.28
(1) isopropanol (2) di-n-propyl-ether (298K)	14	0.32	2.66	-0.32
(1) isopropanol (2) di-n-propyl-ether (313K)	14	0.37	3.08	-0.37
(1) propylene (2) 2-butanol (353K)	12	1.04	10.42	1.04
(1) toluene (2) pentanol (110C)	23	2.47	11.77	-2.43
(1) 5-nonanone (2) 1-hexene (60C)	9	0.49	7.03	-0.07
(1) toluene (2) 1-butanol (90C)	17	1.15	7.69	-1.09
(1) toluene (2)_1-propanol (90C)	17	1.15	7.69	-1.09

From a quick evaluation of the statistics it is clear that the most unsatisfactory results are for the systems involving alcohols, specifically the group interactions C=C/OH and $ACCH_2$/OH. Even though these group interactions were used for regression of the AIPs, it is difficult to know *a priori* for which type of systems these group interactions will give adequate predictions. An important reason is the influence of other systems used for regression that are also dependant on the C-C, and C-O atom interactions (for example, a system involving the group interaction CH_2-CH_2O). Such systems will influence the values of $b_{C\text{-}C}$, $c_{C\text{-}C}$, $e_{C\text{-}C}$, $b_{C\text{-}O}$, $c_{C\text{-}O}$, and $e_{C\text{-}O}$. Generalization in the use of the AIPs implies a strong interdependency between group and atom interactions that is complicated to monitor.

Qualitative Analysis

Several representative phase diagrams of predictions of phase diagrams using GIPs derived from CIs are shown in Figure 5.6.

Figure 5.6a represents the system toluene/1-butanol at 363.15 K, and even though the prediction with the CI-model follows the trend of the experimental data (including the azeotrope) it fails by under-predicting the pressures. Considering the temperature of the system, it is possible that the lack of experimental data covering a broader range of temperatures for aromatic compound-alcohol systems is affecting the accuracy of the prediction.

Two systems involving alcohols and ethers are shown in Figures 5.6b and 5.6c. The systems shown are isopropanol/di-n-propyl-ether and 1-propanol/di-n-propyl-ether, both giving azeotropes. Even though there is a small deviation of predictions, with respect to the experimental data, the systems are represented well with the UNIFAC-CI (Dortmund) model.

For carboxylic acids, the Hayden-O´Connell[81] chemical association theory has been applied whenever needed; but some difficulties were encountered when correlating systems involving carboxylic acids. In Figure 5.6d, the prediction using the CI-model

for the system acetic acid/n-octane at 343.15 K is compared with experimental data - it can be seen that the complexity of the system cannot be fully captured by the CI-model (visually more evident at low compositions of acetic acid), however, the azeotrope is well predicted and the trend seems to be correct regardless of the lack of experimental data at low compositions of acetic acid.

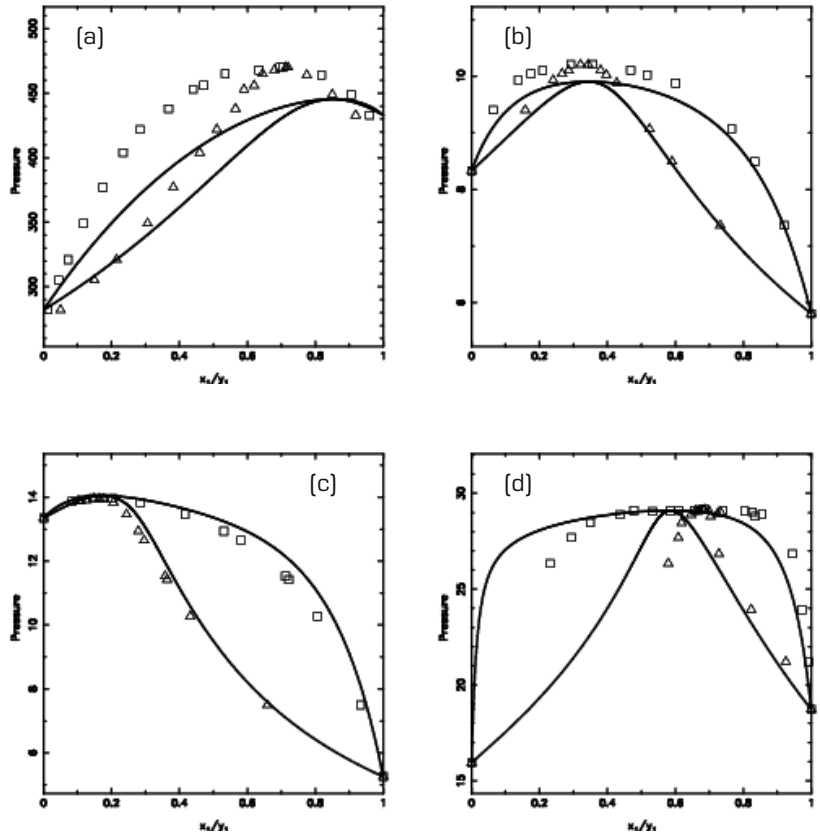


*Figure 5.6 UNIFAC-CI (Dortmund) predictions for the systems: (a) toluene (1) / 1-butanol (2). T = 363.15, P = mmHg. Data taken from: Seetharamaswamy et al., J. App. Chem., **19**, 258 (1969). (b) isopropanol (1) / di-n-propyl-ether (2). T = 298.15 K, P = kPa. Data taken from: Garriga R. et al., Fluid Phase Equilibria, **138**, 131 (1997). (c) 1-propanol (1) / di-n-propyl-ether (2). T = 308.15 K, P(= kPa. Data taken from: Garriga R. et al., Fluid Phase Equilibria, **138**, 131 (1997). (d) acetic acid (1) / n-octane (2). T = 343.15 K, P = kPa. Data taken from: Fu Y., Sandler S., and Orbey H., J. Chem. Eng. Data, **41**, 4 (1996). □-Δ (measured).*

Special cases

Figure 5.7 highlights a case when a prediction is performed for UNIFAC groups not covered in the parameter estimations procedure and that represents a strong extrapolation in the use of AIPs. The system described is n-octane/methyl-methacrylate which involves interactions with the UNIFAC group COO (group 41) that was not used in the regression of AIPs. The CI-derived parameters are not good enough to follow the trend of the experimental pressure, and that is not surprising. The character of this case is exploratory – it gives insight to the reliability of the model when used for extrapolation, and probes the possibility of improvement in the application range of predictions.

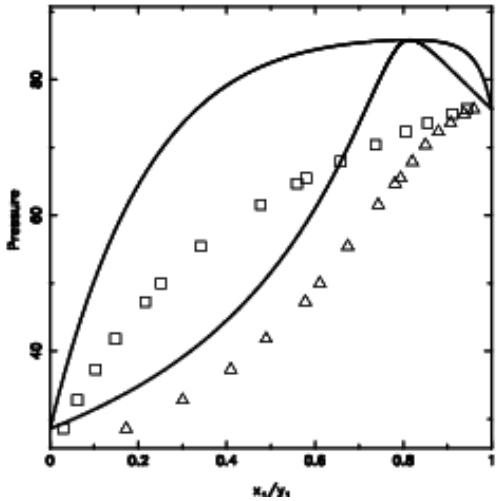


Figure 5.7 UNIFAC-CI (Dortmund) prediction for the system n-hexane (1) – methyl-methacrylate (2). T (=) 333.15 K , P = kPa. Data taken from: Lu B.C., Ishikawa T. and Benson J. Chem. Eng. Data, ***35****, 3 (1990).* □-Δ *(measured).*

Conclusions

Only after proper validation of the UNIFAC-CI (Dortmund) model against more experimental data can this UNIFAC-CI model be used to simulate, analyze, and design industrial processes. A more exhaustive parameterization is needed, as well as the inclusion of more experimental data for correlation, which might allow the addition of more Modified UNIFAC (Dortmund) groups.

5.3 Case Studies

5.3.1 Extrapolation to SLE calculations

Process modeling has not been practiced extensively in the pharmaceutical industry, aside from emission reduction and solvent recovery studies. Recently, the chemical engineering community has identified several areas where process modeling and

molecular thermodynamics could bring significant benefits to the chemicals and fine chemicals industries. For example, in the pharmaceutical industry, when screening for drug candidates, it is very important to develop process recipes that involve multiple reaction steps and separation steps, such as, crystallization or extraction. Formulation chemists are interested in understanding and enhancing drug solubility in therapeutical formulations. The industry is interested in solubilities for millions of drug-like molecules at physiological conditions in the compound libraries that they screen in search of good molecules.

It is well known that the use of the UNIFAC models is not restricted to VLE calculations; historically the UNIFAC models have been also used for SLE calculation: Gmehling et al.[73] showed moderately good results extrapolating the GIPs derived mainly from VLE data for Original UNIFAC. Also, the use of more sophisticated UNIFAC models[73, 19, 1], with GIPs based on VLE, SLE, LLE, and H^{EX} data (among other), have made possible the qualitative prediction of SLE in a successful manner.

However, the limited experimental resources and compound availability sometimes reflect the inadequacy of the UNIFAC models for solubility modeling due to gaps in the group interaction parameter tables, missing functional groups and the collapse of the functional group additivity rule with large, complex drug-like molecules. Whenever possible, the UNIFAC-CI models can help overcome these situations by extending the application range of the UNIFAC models for solubility calculations. The purpose of this section is to study the possibilities of using the UNIFAC GIPs generated through CIs for the 'prediction' of SLE calculations. This is illustrated through several examples, in which the CI-derived GIPs were used to compare the qualitative performance of the UNIFAC-CI model against experimental data involving important solvents for the chemical industry.

In Figure 5.8, the GIPs generated through CIs for UNIFAC-CI (VLE) were used extrapolatively to 'predict' the SLE of naphthalene with: a) acetic acid (ACH-CH_2, ACH-COOH group interactions); b) acetone (ACH-CH_2, ACH-CH_2CO group interactions); c) n-hexane (ACH-CH_2 group interactions); and d) 1-butanol (ACH-CH_2, ACH-OH group interactions). For all the cases, the agreement with experimental data is considered to be good, however, some inaccuracies are observed for the system with acetic acid (Figure 5.8a). In general, it seems that the same strengths and limitations expressed in UNIFAC (VLE) are transferred to SLE calculations.

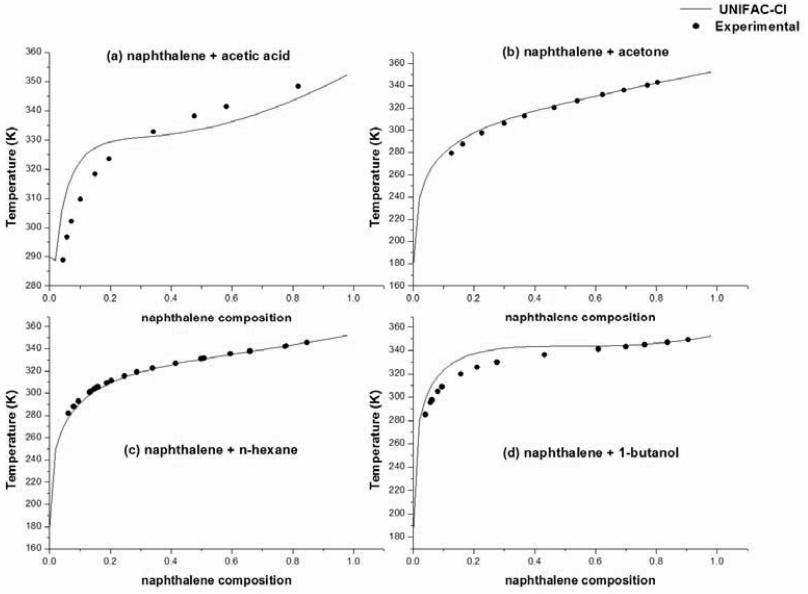


Figure 5.8 UNIFAC-CI SLE predictions (using full CI-generated GIPs based on VLE data). T (=) K. Data taken from: Ward L.H., J. Phys. Chem., 30, 10 (1926). ● (measured).

Figure 5.9 highlights the SLE predictions for the following systems: a) 1,4- xylene/1,3 xylene (CH_2-ACH group interactions); b) acetic acid/benzene (CH_2-ACH, ACH-COOH); c) phenol/ethanol (CH_2-ACH, CH_2-OH, CH_2-ACOH, ACH-OH, ACH-ACOH, OH-ACOH group interactions); and d) phenol/benzene (ACH-ACOH group interactions). In Figure 5.9a, an excellent agreement with experimental data can be observed, which was expected due to the high ideality of the system, while for the system in Figure 5.9b there is a pronounced displacement of the predicted eutectic point compared to experimental data. The same happens for the system in Figure 5.9d, but the difference between the experimental eutectic point and the predicted one is with respect to temperature. These systems show moderate deviations when predicting their VLE using UNIFAC-CI which is the cause for the inability to capture the entire trend of the experimental data. It should not be forgotten however that the used GIPs have been derived from a limited amount of experimental VLE data. Finally, in Figure 5.9c, the prediction for phenol/ethanol system deviates more as the phenol composition decreases, and strictly speaking, the experimental data is incomplete and therefore it is not possible to assess the quality of the prediction below phenol composition value of 0.5.

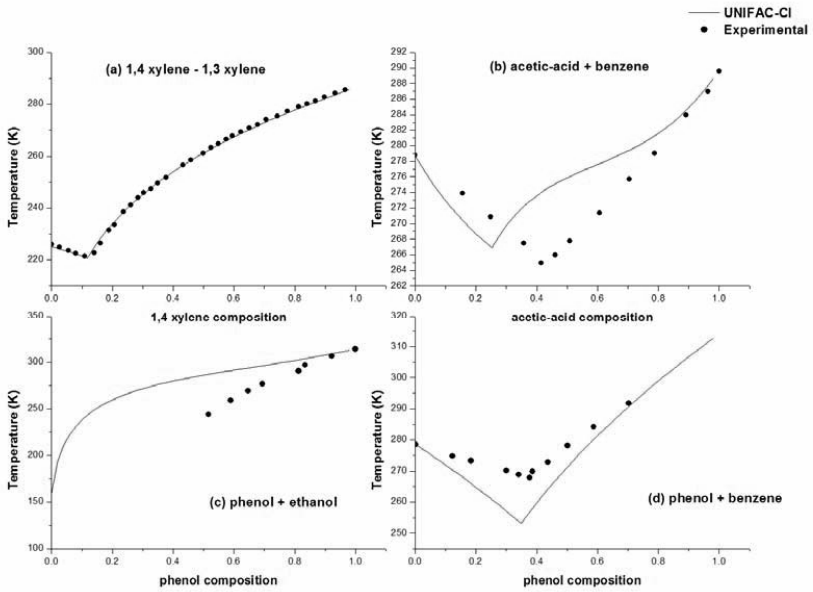


*Figure 5.9 UNIFAC-CI SLE predictions using (using full CI-generated GIPs from VLE data). Data taken from: (a) Jakob A. et al., Fluid Phase Equilibria, 113 (1996). (b and d) Gmehling et al., Ind. Eng. Chem. Fundam., **17**, 4 (1978). (c) Perrakis, N., J. Chim. Phys., **22**, 280 (1925). ● (measured). T = K.*

Finally, the SLE predictions for the systems benzene/pyridine, indane/dodecane, and indane/1,2-dichloroethane are shown in Figure 5.10. All the predictions using the CI-derived GIPs seem to follow the same pattern as experimental data. In Figure 5.10a, benzene/pyridine (involving the interactions ACH-PYR) system is shown; the CI-model follows correctly the experimental data except for a small discrepancy for the eutectic point at benzene molar composition value of about 0.4. The system indane/dodecane (group interactions CH_2-ACH, CH2-$ACCH_2$, ACH-$ACCH_2$) has an excellent agreement with experimental data which was expected considering the group interactions involved which have been parameterized from a broad experimental data set. The system indane/1-2 dichloroethane (group interactions CH_2-ACH, CH_2-$ACCH_2$, CH_2-CCl, ACH-$ACCH_2$, ACH-CCl, and $ACCH_2$-CCl), shown in Figure 5.10c, represents a more 'complicated' system involving four main UNIFAC groups and a heteroatom (chlorine). Good agreement was obtained.

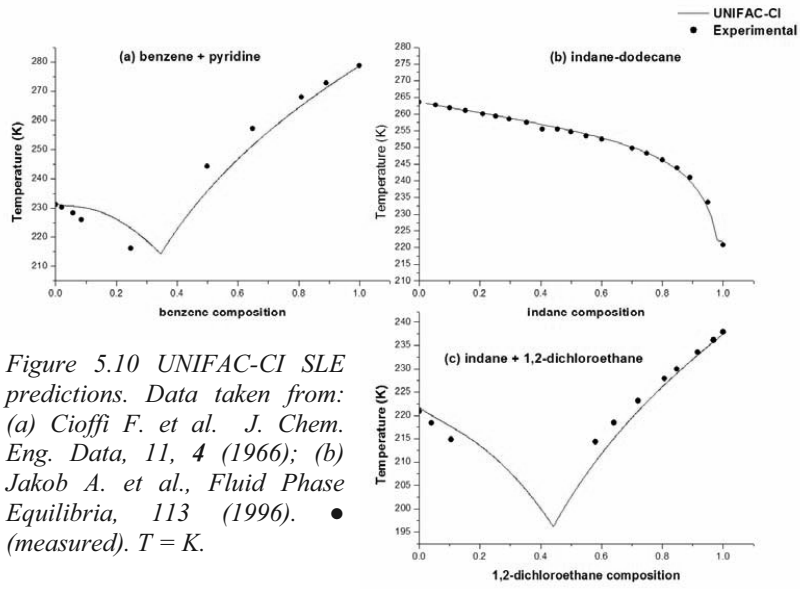


Figure 5.10 UNIFAC-CI SLE predictions. Data taken from: (a) Cioffi F. et al. J. Chem. Eng. Data, 11, 4 (1966); (b) Jakob A. et al., Fluid Phase Equilibria, 113 (1996). ● *(measured). T = K.*

The results above, though not very extensive, are encouraging and promising for limited predictions using UNIFAC for SLE calculations. More investigation is however needed, with the aim of establishing the limits in the application for the UNIFAC-CI models to SLE calculations. Also, the generation of new functional groups suitable for prediction of solvent properties using the rules of the CI methodology is very attractive. This can be considered for the process design, and combined process and product design of chemical substances that at present cannot be described using UNIFAC.

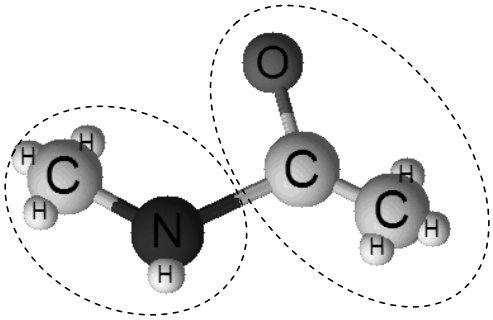
As mentioned above, the same application range regarding the chemical species is expected for SLE calculations as for original UNIFAC-CI (VLE). Since no SLE data was used for the generation of the AIPs, the extrapolation to SLE calculations should be done with caution and rough prediction accuracy can be expected.

5.3.2 Use of the CI-based GIPs: Filling the Gaps in the Original UNIFAC matrix

Pyridine/n-methylacetamide case

One of the main objectives when applying GC^{Plus} methodology to the UNIFAC models, is getting the possibility of extending the range of application by enabling calculations that were not feasible previously because of the missing GIPs. The UNIFAC-CI method provides the possibility of filling the gaps in the parameter table. In this case study, the prediction of a system using CI-generated GIPs which were not previously available is highlighted.

The system under study is pyridine/n-methylacetamide. n-Methylacetamide has been described here by using the following UNIFAC groups: CH_2CO (main group 9), and CNH (main group 15). The UNIFAC description of n-methylacetamide is illustrated in Figure 5.10.



1 x CNH + 1 x CH_2CO

Figure 5.11 Representation of the n-methylacetamide in terms of UNIFAC groups, using the CH_2CO and CNH main groups.

Please note that n-methylacetamide can also be represented by using the UNIFAC main groups CH_2 and CON. Since the objective here however, is to highlight how a missing GIP of the reference UNIFAC model[71] is created and tested for verification, the CON (main group 46) based representation has been discarded because that representation does not have any gaps in the UNIFAC parameter table.

For the system under study, making an inspection of the available parameter table for the reference UNIFAC version[71] (see Table 5.3), shows that a couple of GIPs are missing (highlighted in Table 5.3). Those are corresponding to the CNH-PYRIDINE and PYRIDINE-CNH group interactions.

Table 5.3 UNIFAC parameter table for the system pyridine/n-methylacetamide

	CH2CO	CNH	PYRIDINE
CH2CO	0.0	-174.2	7.341
CNH	394.6	0.0	n.a.
PYRIDINE	29.1	n.a.	0.0

Therefore, the use of UNIFAC-CI provides the following possibilities in order to allow phase equilibria calculations using UNIFAC:

- Approach A -Use CIs to generate only the GIPs that are missing, while using original values for toher GIPs;
- Approach B - Use the GIPs generated through CIs for all the group interactions.

The values for a_{mn} and a_{nm} (where m=CNH and n=PYRIDINE) are obtained in the following way:

1) Since PYRIDINE is a molecular group, the special table is needed to generate the GIPs related to PYRIDINE (see Appendix B, Table B.4). All the other interactions are obtained from the 'functional' AIPs table (see Appendix B, Table B.1);
2) The stoichiometry and the CIs values needed for the calculation are summarized in Table 5.4;

Table 5.4 CI and Stoichiometric Values for the UNIFAC Groups CNH and PYRIDINE

	n_C	n_O	n_N	${}^0\chi^v$	${}^1\chi^v$	${}^2\chi^v$
CNH	1	0	1	1.50000	0.500000	0.000000
PYRIDINE	5	0	1	1.98289	0.655262	0.218420

3) Using the expression derived in Equation (3.7), the CNH-PYRINIDE GIPs are calculated as follows:

$$a_{CNH-PYR} = b_{C-C}\frac{n_C^{(CNH)}\chi_0^{(PYR)} - n_C^{(PYR)}\chi_0^{(CNH)}}{\chi_0^{(PYR)}\chi_0^{(CNH)}} + c_{C-C}\frac{n_C^{(CNH)}\chi_1^{(PYR)} - n_C^{(PYR)}\chi_0^{(CNH)}}{\chi_1^{(PYR)}\chi_0^{(CNH)}}$$
$$+ d_{C-C}\frac{n_C^{(CNH)}\chi_1^{(PYR)} - n_C^{(PYR)}\chi_1^{(CNH)}}{\chi_1^{(PYR)}\chi_1^{(CNH)}} + e_{C-C}\frac{n_C^{(CNH)}\chi_2^{(PYR)} - n_C^{(PYR)}\chi_0^{(CNH)}}{\chi_2^{(PYR)}\chi_0^{(CNH)}}$$

$$+ b_{C-N}\frac{n_C^{(CNH)}\chi_0^{(PYR)} - n_N^{(PYR)}\chi_0^{(CNH)}}{\chi_0^{(PYR)}\chi_0^{(CNH)}} + c_{C-N}\frac{n_C^{(CNH)}\chi_1^{(PYR)} - n_N^{(PYR)}\chi_0^{(CNH)}}{\chi_1^{(PYR)}\chi_0^{(CNH)}}$$
$$+ d_{C-N}\frac{n_C^{(CNH)}\chi_1^{(PYR)} - n_N^{(PYR)}\chi_1^{(CNH)}}{\chi_1^{(PYR)}\chi_1^{(CNH)}} + e_{C-N}\frac{n_C^{(CNH)}\chi_2^{(PYR)} - n_N^{(PYR)}\chi_0^{(CNH)}}{\chi_2^{(PYR)}\chi_0^{(CNH)}}$$

$$+ b_{N-C}\frac{n_N^{(CNH)}\chi_0^{(PYR)} - n_C^{(PYR)}\chi_0^{(CNH)}}{\chi_0^{(PYR)}\chi_0^{(CNH)}} + c_{N-C}\frac{n_N^{(CNH)}\chi_1^{(PYR)} - n_C^{(PYR)}\chi_0^{(CNH)}}{\chi_1^{(PYR)}\chi_0^{(CNH)}}$$
$$+ d_{N-C}\frac{n_N^{(CNH)}\chi_1^{(PYR)} - n_C^{(PYR)}\chi_1^{(CNH)}}{\chi_1^{(PYR)}\chi_1^{(CNH)}} + e_{N-C}\frac{n_N^{(CNH)}\chi_2^{(PYR)} - n_C^{(PYR)}\chi_0^{(CNH)}}{\chi_2^{(PYR)}\chi_0^{(CNH)}}$$

$$+ b_{N-N}\frac{n_N^{(CNH)}\chi_0^{(PYR)} - n_N^{(PYR)}\chi_0^{(CNH)}}{\chi_0^{(PYR)}\chi_0^{(CNH)}} + c_{N-N}\frac{n_N^{(CNH)}\chi_1^{(PYR)} - n_N^{(PYR)}\chi_0^{(CNH)}}{\chi_1^{(PYR)}\chi_0^{(CNH)}}$$
$$+ d_{N-N}\frac{n_N^{(CNH)}\chi_1^{(PYR)} - n_O^{(PYR)}\chi_1^{(CNH)}}{\chi_1^{(PYR)}\chi_1^{(CNH)}} + e_{N-N}\frac{n_N^{(CNH)}\chi_2^{(PYR)} - n_O^{(PYR)}\chi_0^{(CNH)}}{\chi_2^{(PYR)}\chi_0^{(CNH)}}$$

The needed AIPs to generate the $\boldsymbol{a}_{CNH\text{-}PYR}$ group interaction parameter are given in Table 5.5.

Table 5.5 Atom Interaction Parameters needed for the calculation of the Group Interaction Parameter $a_{CNH\text{-}PYR}$

$b_{C\text{-}C}$= 988.0659	$c_{C\text{-}C}$= -84.2139	$d_{C\text{-}C}$= 28.9377	$e_{C\text{-}C}$= -44.8237
$b_{C\text{-}N}$= -402.1131	$c_{C\text{-}N}$= -77.3075	$d_{C\text{-}N}$= 3590.7041	$e_{C\text{-}N}$= -98.8721
$b_{N\text{-}C}$= -306.9166	$c_{N\text{-}C}$= 120.5933	$d_{N\text{-}C}$= -1632.5328	$e_{N\text{-}C}$= 467.2085
$b_{N\text{-}N}$= -401.3424	$c_{N\text{-}N}$= -170.5513	$d_{N\text{-}N}$= -1921.2043	$e_{N\text{-}N}$= -79.9868

When substituting all the needed AIPs, number of atoms, and CIs values, the obtained CNH-PYRIDINE group interaction parameter is $a_{CNH\text{-}PYR}$= -301.81.

The mirror parameter $a_{PYR\text{-}CNH}$ is calculated following the same procedure. The reduced form of Equation (3.8) is the following:

$$a_{PYR-CNH} = \overline{b_{C-C}} \frac{n_C^{(PYR)}\chi_0^{(CNH)} - n_C^{(CNH)}\chi_0^{(PYR)}}{\chi_0^{(CNH)}\chi_0^{(PYR)}} + \overline{c_{C-C}} \frac{n_C^{(PYR)}\chi_1^{(CNH)} - n_C^{(CNH)}\chi_0^{(PYR)}}{\chi_1^{(CNH)}\chi_0^{(PYR)}}$$
$$+ \overline{d_{C-C}} \frac{n_C^{(PYR)}\chi_1^{(CNH)} - n_C^{(CNH)}\chi_1^{(PYR)}}{\chi_1^{(CNH)}\chi_1^{(PYR)}} + \overline{b_{C-N}} \frac{n_C^{(PYR)}\chi_0^{(CNH)} - n_N^{(CNH)}\chi_0^{(PYR)}}{\chi_0^{(CNH)}\chi_0^{(PYR)}}$$

$$+ \overline{c_{C-N}} \frac{n_C^{(PYR)}\chi_1^{(CNH)} - n_N^{(CNH)}\chi_0^{(PYR)}}{\chi_1^{(CNH)}\chi_0^{(PYR)}} + \overline{d_{C-N}} \frac{n_C^{(PYR)}\chi_1^{(CNH)} - n_N^{(CNH)}\chi_1^{(PYR)}}{\chi_1^{(CNH)}\chi_1^{(PYR)}}$$
$$+ \overline{b_{N-C}} \frac{n_N^{(PYR)}\chi_0^{(CNH)} - n_C^{(CNH)}\chi_0^{(PYR)}}{\chi_0^{(CNH)}\chi_0^{(PYR)}} + \overline{c_{N-C}} \frac{n_N^{(PYR)}\chi_1^{(CNH)} - n_C^{(CNH)}\chi_0^{(PYR)}}{\chi_1^{(CNH)}\chi_0^{(PYR)}}$$

$$+ \overline{d_{N-C}} \frac{n_N^{(PYR)}\chi_1^{(CNH)} - n_C^{(CNH)}\chi_1^{(PYR)}}{\chi_1^{(CNH)}\chi_1^{(PYR)}} + \overline{b_{N-N}} \frac{n_N^{(PYR)}\chi_0^{(CNH)} - n_N^{(CNH)}\chi_0^{(PYR)}}{\chi_0^{(CNH)}\chi_0^{(PYR)}}$$
$$+ \overline{c_{N-N}} \frac{n_N^{(PYR)}\chi_1^{(CNH)} - n_N^{(CNH)}\chi_0^{(PYR)}}{\chi_1^{(CNH)}\chi_0^{(PYR)}} + \overline{d_{N-N}} \frac{n_N^{(PYR)}\chi_1^{(CNH)} - n_O^{(CNH)}\chi_1^{(PYR)}}{\chi_1^{(CNH)}\chi_1^{(PYR)}}$$

The AIPs needed to generate the $a_{PYR\text{-}CNH}$ group interaction parameter are given in Table 5.6.

Table 5.6 Atom Interaction Parameters needed for the calculation of the Group Interaction Parameter $a_{CNH\text{-}PYR}$

$bh_{C\text{-}C}$= -3.4234	$ch_{C\text{-}C}$= -242.2328	$dh_{C\text{-}C}$= 402.4844
$bh_{C\text{-}N}$= -171.1524	$ch_{C\text{-}N}$= 1289.0657	$dh_{C\text{-}N}$= -69.0159
$bh_{N\text{-}C}$= 116.7745	$ch_{N\text{-}C}$= -1495.5216	$dh_{N\text{-}C}$= 2222.4245
$bh_{N\text{-}N}$= 114.7603	$ch_{N\text{-}N}$= 1577.9300	$dh_{N\text{-}N}$= 831.1490

Finally, when substituting all the needed AIPs, number of atoms and CIs values, the obtained value of the CNH-PYRIDINE group interaction parameter is $\boldsymbol{a}_{PYR\text{-}CNH}$ = 491.85.

The missing GIPs have now been 'generated' through the UNIFAC-CI methodology. The experimental verification, can be seen in Figure 5.12 where there is a comparison of the experimental data for the system pyridine/n-methylacetamide with the performance of UNIFAC-CI using full CI-generated GIPs (approach B), and the performance of UNIFAC-CI where substitution is made just for the pair of GIPs missing in the parameter table (approach A).

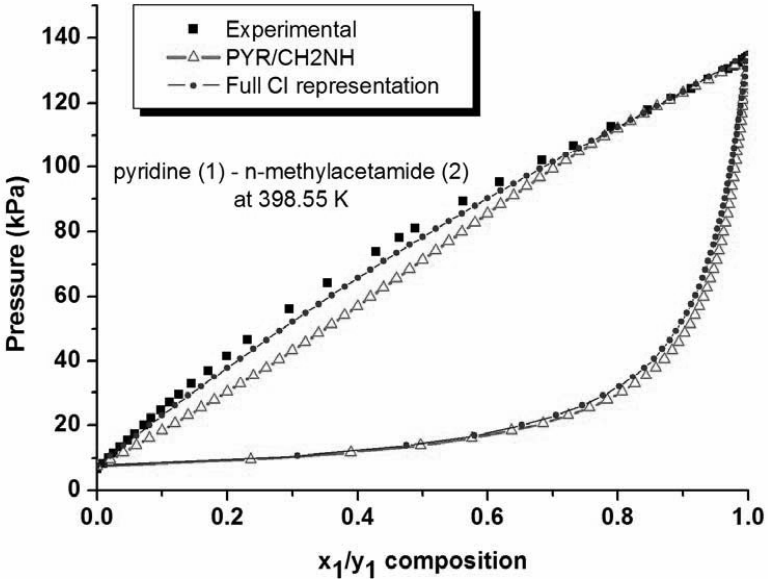


Figure 5.12 UNIFAC-CI model prediction for the system pyridine (1) / n-methylacetamide (2). T = 398.15. Comparison between using full CI-generated GIPs (-●-) and using CI-approach just for the missing pair CNH-PYR (Δ). Data taken from: Haan, A.B. et al. J. Chem. Eng. Data, 42, 597-602 (1997).

The UNIFAC parameter table extracts used are summarized in Tables 5.7a and 5.7b:

Table 5.7a Parameter table including the pair for the CNH-PYR interaction

	CH_2CO	CNH	PYRIDINE
CH_2CO	0.0	-174.2	7.341
CNH	394.6	0.0	-301.81
PYRIDINE	29.1	491.85	0.0

Table 5.7.b Parameter table with full CI-generated GIPs.

	CH_2CO	CNH	PYRIDINE
CH_2CO	0.0	3650.45	12830.41
CNH	-291.79	0.0	-301.81
PYRIDINE	-302.46	491.85	0.0

In Figure 5.12, it can be seen that the two approaches give good representation of the experimental data. The main discrepancies between the approaches occur below molar composition value of 0.5 for pyridine where the full CI-representation gives slightly better results. An alternative for prediction when UNIFAC GIPs are not available has been presented, and the approaches A and B fulfill part of the objectives of this work, for the case study that has been described. However, the recommendation in general is for the user to limit the use of CIs just to GIPs missing in the UNIFAC parameter table. An important reason is that the parameterization of UNIFAC has been carefully performed for more than 30 years which gives its well known robustness and reliability. That is why the original GIPs should be used and the gaps should be left for the CI-method. For assesment of the reliability when coupling CI-based GIPs with the original GIPs, predictions for more systems are needed (especially highly non-ideal systems), as well as a significant statistical comparison of the approaches A & B for a number of

different systems. Some other systems with GIPs missing in the original UNIFAC matrix have been shown previously[90, 91] in other works with encouraging results.

System 1,2-dichloroethane/dimethyl-sulfoxide (DMSO)

DMSO is a dipolar, aprotic solvent that is considered very important in the chemical industry. Due to its solvating properties it is popular as a solvent to promote chemical reactions. It is a highly polar liquid and has the ability to participate in hydrogen bonding. Also, the interactions of highly polar compounds (such as DMSO) with chloroethanes have recently received attention because they serve as model compounds in biochemical processes[92]. An important amount of UNIFAC GIPs involving DMSO and chlorinated compounds is however unavailable (according to the reference UNIFAC model[71] used in this work). That is the reason why this system has been chosen for a case study to evaluate the performance of the predictions when using GIPs involving 1,2-dichloroethane and dimethyl-sulfoxide.

The representation of both compounds through UNIFAC groups is shown in Figure 5.13.

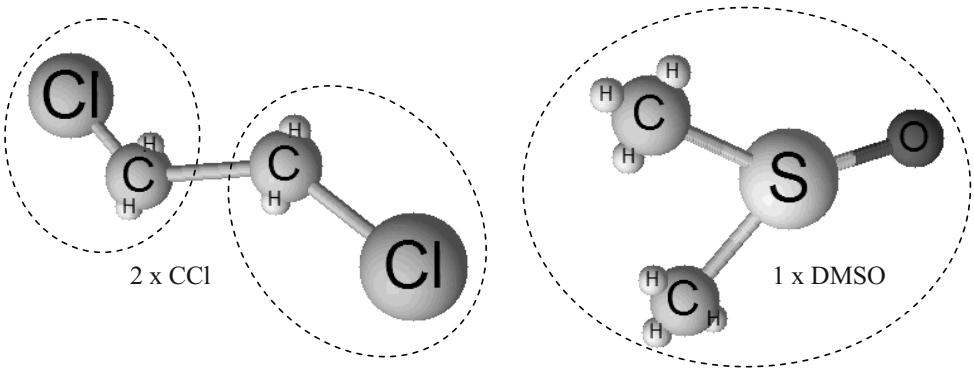


Figure 5.13 Representation through UNIFAC main groups of the compounds 1,2-dichloroethane (left) and DMSO (right)

The needed matrix of GIPs therefore consists of a 2 x 2 matrix involving the UNIFAC groups CCl (main group 21) and DMSO (main group 35). In the reference model, the GIPs for the interaction CCl-DMSO (and mirror) are not available (as it can be seen in Table 5.8). Therefore the estimated AIPs will be used to calculate this missing interaction. The AIPs are extracted from the parameter Table B.7, (see Appendix B).

Table 5.8 UNIFAC parameter table for the system 1,2-dichloroethane/DMSO in the reference UNIFAC models (original UNIFAC[71])

	CCl	DMSO
CCl	0.0	n.a.
DMSO	n.a.	0.0

The values of a_{mn} and a_{nm} (where m=CCl and n=DMSO) are obtained in the following way:

1) DMSO is a molecular group, and a special table is needed to generate the GIPs involving DMSO (see Appendix B, table B.7). There are no other UNIFAC group interactions involved.

2) The stoichiometry and the CI values needed for the interaction, are summarized in Table 5.9.

Table 5.9 CI and Stoichiometric Values for the UNIFAC Groups CCl and DMSO

	n_C	n_O	n_N	n_{Cl}	n_S	${}^0\chi^v$	${}^1\chi^v$	${}^2\chi^v$
CCl	1	0	0	1	0	2.133893	1.133893	0.000000
DMSO	2	1	0	0	1	2.816496	0.983163	0.741581

3) Using the expression derived in Equation 3.7, the CCl-DMSO GIPs are calculated as follows:

$$a_{CCl-DMSO} = b_{C-C}\frac{n_C^{(CCl)}\chi_0^{(DMSO)} - n_C^{(DMSO)}\chi_0^{(CCl)}}{\chi_0^{(DMSO)}\chi_0^{(CCl)}} + c_{C-C}\frac{n_C^{(CCl)}\chi_1^{(DMSO)} - n_C^{(DMSO)}\chi_0^{(CCl)}}{\chi_1^{(DMSO)}\chi_0^{(CCl)}}$$
$$+ d_{C-C}\frac{n_C^{(CCl)}\chi_1^{(DMSO)} - n_C^{(DMSO)}\chi_1^{(CCl)}}{\chi_1^{(DMSO)}\chi_1^{(CCl)}} + e_{C-C}\frac{n_C^{(CCl)}\chi_2^{(DMSO)} - n_C^{(DMSO)}\chi_0^{(CCl)}}{\chi_2^{(DMSO)}\chi_0^{(CCl)}}$$

$$+ b_{C-O}\frac{n_C^{(CCl)}\chi_0^{(DMSO)} - n_O^{(DMSO)}\chi_0^{(CCl)}}{\chi_0^{(DMSO)}\chi_0^{(CCl)}} + c_{C-O}\frac{n_C^{(CCl)}\chi_1^{(DMSO)} - n_O^{(DMSO)}\chi_0^{(CCl)}}{\chi_1^{(DMSO)}\chi_0^{(CCl)}}$$
$$+ d_{C-O}\frac{n_C^{(CCl)}\chi_1^{(DMSO)} - n_O^{(DMSO)}\chi_1^{(CCl)}}{\chi_1^{(DMSO)}\chi_1^{(CCl)}} + e_{C-O}\frac{n_C^{(CCl)}\chi_2^{(DMSO)} - n_O^{(DMSO)}\chi_0^{(CCl)}}{\chi_2^{(DMSO)}\chi_0^{(CCl)}}$$

$$+ b_{C-S}\frac{n_C^{(CCl)}\chi_0^{(DMSO)} - n_S^{(DMSO)}\chi_0^{(CCl)}}{\chi_0^{(DMSO)}\chi_0^{(CCl)}} + c_{C-S}\frac{n_C^{(CCl)}\chi_1^{(DMSO)} - n_S^{(DMSO)}\chi_0^{(CCl)}}{\chi_1^{(DMSO)}\chi_0^{(CCl)}}$$
$$+ d_{C-S}\frac{n_C^{(CCl)}\chi_1^{(DMSO)} - n_S^{(DMSO)}\chi_1^{(CCl)}}{\chi_1^{(DMSO)}\chi_1^{(CCl)}} + e_{C-S}\frac{n_C^{(CCl)}\chi_2^{(DMSO)} - n_S^{(DMSO)}\chi_0^{(CCl)}}{\chi_2^{(DMSO)}\chi_0^{(CCl)}}$$

$$+ b_{Cl-C}\frac{n_{Cl}^{(CCl)}\chi_0^{(DMSO)} - n_C^{(DMSO)}\chi_0^{(CCl)}}{\chi_0^{(DMSO)}\chi_0^{(CCl)}} + c_{Cl-C}\frac{n_{Cl}^{(CCl)}\chi_1^{(DMSO)} - n_C^{(DMSO)}\chi_0^{(CCl)}}{\chi_1^{(DMSO)}\chi_0^{(CCl)}}$$
$$+ d_{Cl-C}\frac{n_{Cl}^{(CCl)}\chi_1^{(DMSO)} - n_C^{(DMSO)}\chi_1^{(CCl)}}{\chi_1^{(DMSO)}\chi_1^{(CCl)}} + e_{Cl-C}\frac{n_{Cl}^{(CCl)}\chi_2^{(DMSO)} - n_C^{(DMSO)}\chi_0^{(CCl)}}{\chi_2^{(DMSO)}\chi_0^{(CCl)}}$$

$$+ b_{Cl-O}\frac{n_{Cl}^{(CCl)}\chi_0^{(DMSO)} - n_O^{(DMSO)}\chi_0^{(CCl)}}{\chi_0^{(DMSO)}\chi_0^{(CCl)}} + c_{Cl-O}\frac{n_{Cl}^{(CCl)}\chi_1^{(DMSO)} - n_O^{(DMSO)}\chi_0^{(CCl)}}{\chi_1^{(DMSO)}\chi_0^{(CCl)}}$$
$$+ d_{Cl-O}\frac{n_{Cl}^{(CCl)}\chi_1^{(DMSO)} - n_O^{(DMSO)}\chi_1^{(CCl)}}{\chi_1^{(DMSO)}\chi_1^{(CCl)}} + e_{Cl-O}\frac{n_{Cl}^{(CCl)}\chi_2^{(DMSO)} - n_O^{(DMSO)}\chi_0^{(CCl)}}{\chi_2^{(DMSO)}\chi_0^{(CCl)}}$$

$$+ b_{Cl-S}\frac{n_{Cl}^{(CCl)}\chi_0^{(DMSO)} - n_S^{(DMSO)}\chi_0^{(CCl)}}{\chi_0^{(DMSO)}\chi_0^{(CCl)}} + c_{Cl-S}\frac{n_{Cl}^{(CCl)}\chi_1^{(DMSO)} - n_S^{(DMSO)}\chi_0^{(CCl)}}{\chi_1^{(DMSO)}\chi_0^{(CCl)}}$$
$$+ d_{Cl-S}\frac{n_{Cl}^{(CCl)}\chi_1^{(DMSO)} - n_S^{(DMSO)}\chi_1^{(CCl)}}{\chi_1^{(DMSO)}\chi_1^{(CCl)}} + e_{Cl-S}\frac{n_{Cl}^{(CCl)}\chi_2^{(DMSO)} - n_S^{(DMSO)}\chi_0^{(CCl)}}{\chi_2^{(DMSO)}\chi_0^{(CCl)}}$$

The AIPs needed to generate the $\boldsymbol{a}_{CCl\text{-}DMSO}$ group interaction parameter are given in Table 5.10.

Table 5.10 Atom Interaction Parameters needed for the calculation of the Group Interaction Parameter $a_{CCl\text{-}DMSO}$

$b_{C\text{-}C}$= 975.4933	$b_{C\text{-}O}$= -1136.644	$b_{C\text{-}S}$=378.9446	$b_{Cl\text{-}C}$=-0.3767	$b_{Cl\text{-}O}$=0.7864	$b_{Cl\text{-}S}$=61.0494
$b_{C\text{-}C}$= -112.1958	$c_{C\text{-}O}$= -20.1612	$c_{C\text{-}S}$=842.4080	$c_{Cl\text{-}C}$=-4.7144	$c_{Cl\text{-}O}$=-1.3824	$c_{Cl\text{-}S}$=-794.4491
$d_{C\text{-}C}$=100.8302	$d_{C\text{-}O}$= 44.9423	$d_{C\text{-}S}$=-732.2557	$d_{Cl\text{-}C}$=-3.6198	$d_{Cl\text{-}O}$=-0.2879	$d_{Cl\text{-}S}$=-224.0920
$b_{C\text{-}C}$= -114.4053	$e_{C\text{-}O}$= 58.6682	$e_{C\text{-}S}$=195.6386	$e_{Cl\text{-}C}$=-6.8854	$e_{Cl\text{-}O}$=-2.4678	$e_{Cl\text{-}S}$=63.3785

Substituting all of the AIPs, the number of atoms, and the CIs values, the resulting value of the CCl-DMSO group interaction parameter is $a_{CCl\text{-}DMSO}$= -141.77.

The mirror parameter $a_{DMSO\text{-}CCl}$ is calculated following the same procedure. The reduced form of Equation (3.8) is the following:

$$a_{DMSO-CCl} = \overline{b_{C-C}}\frac{n_C^{(DMSO)}\chi_0^{(CCl)} - n_C^{(CCl)}\chi_0^{(DMSO)}}{\chi_0^{(CCl)}\chi_0^{(DMSO)}} + \overline{c_{C-C}}\frac{n_C^{(DMSO)}\chi_1^{(CCl)} - n_C^{(CCl)}\chi_0^{(DMSO)}}{\chi_1^{(CCl)}\chi_0^{(DMSO)}}$$
$$+ \overline{d_{C-C}}\frac{n_C^{(DMSO)}\chi_1^{(CCl)} - n_C^{(CCl)}\chi_1^{(DMSO)}}{\chi_1^{(CCl)}\chi_1^{(DMSO)}} + \overline{b_{C-Cl}}\frac{n_C^{(DMSO)}\chi_0^{(CCl)} - n_{Cl}^{(CCl)}\chi_0^{(DMSO)}}{\chi_0^{(CCl)}\chi_0^{(DMSO)}}$$

$$+ \overline{c_{C-Cl}}\frac{n_C^{(DMSO)}\chi_1^{(CCl)} - n_{Cl}^{(CCl)}\chi_0^{(DMSO)}}{\chi_1^{(CCl)}\chi_0^{(DMSO)}} + \overline{d_{C-Cl}}\frac{n_C^{(DMSO)}\chi_1^{(CCl)} - n_{Cl}^{(CCl)}\chi_1^{(DMSO)}}{\chi_1^{(CCl)}\chi_1^{(DMSO)}}$$
$$+ \overline{b_{O-C}}\frac{n_O^{(DMSO)}\chi_0^{(CCl)} - n_C^{(CCl)}\chi_0^{(DMSO)}}{\chi_0^{(CCl)}\chi_0^{(DMSO)}} + \overline{c_{O-C}}\frac{n_O^{(DMSO)}\chi_1^{(CCl)} - n_C^{(CCl)}\chi_0^{(DMSO)}}{\chi_1^{(CCl)}\chi_0^{(DMSO)}}$$

$$+ \overline{d_{O-C}}\frac{n_O^{(DMSO)}\chi_1^{(CCl)} - n_C^{(CCl)}\chi_1^{(DMSO)}}{\chi_1^{(CCl)}\chi_1^{(DMSO)}} + \overline{b_{O-Cl}}\frac{n_O^{(DMSO)}\chi_0^{(CCl)} - n_{Cl}^{(CCl)}\chi_0^{(DMSO)}}{\chi_0^{(CCl)}\chi_0^{(DMSO)}}$$
$$+ \overline{c_{O-Cl}}\frac{n_O^{(DMSO)}\chi_1^{(CCl)} - n_{Cl}^{(CCl)}\chi_0^{(DMSO)}}{\chi_1^{(CCl)}\chi_0^{(DMSO)}} + \overline{d_{O-Cl}}\frac{n_O^{(DMSO)}\chi_1^{(CCl)} - n_{Cl}^{(CCl)}\chi_1^{(DMSO)}}{\chi_1^{(CCl)}\chi_1^{(DMSO)}}$$

$$+ \overline{b_{S-C}}\frac{n_S^{(DMSO)}\chi_0^{(CCl)} - n_C^{(CCl)}\chi_0^{(DMSO)}}{\chi_0^{(CCl)}\chi_0^{(DMSO)}} + \overline{c_{S-C}}\frac{n_S^{(DMSO)}\chi_1^{(CCl)} - n_C^{(CCl)}\chi_0^{(DMSO)}}{\chi_1^{(CCl)}\chi_0^{(DMSO)}}$$
$$+ \overline{d_{S-C}}\frac{n_S^{(DMSO)}\chi_1^{(CCl)} - n_C^{(CCl)}\chi_1^{(DMSO)}}{\chi_1^{(CCl)}\chi_1^{(DMSO)}} + \overline{b_{S-Cl}}\frac{n_S^{(DMSO)}\chi_0^{(CCl)} - n_{Cl}^{(CCl)}\chi_0^{(DMSO)}}{\chi_0^{(CCl)}\chi_0^{(DMSO)}}$$

$$+ \overline{c_{S-Cl}}\frac{n_S^{(DMSO)}\chi_1^{(CCl)} - n_{Cl}^{(CCl)}\chi_0^{(DMSO)}}{\chi_1^{(CCl)}\chi_0^{(DMSO)}} + \overline{d_{S-Cl}}\frac{n_S^{(DMSO)}\chi_1^{(CCl)} - n_{Cl}^{(CCl)}\chi_1^{(DMSO)}}{\chi_1^{(CCl)}\chi_1^{(DMSO)}}$$

The AIPs needed to generate the $a_{DMSO\text{-}CCl}$ group interaction parameter are given in Table 5.11.

Table 5.11 Atom Interaction Parameters needed for the calculation of the Group Interaction Parameter $a_{DMSO\text{-}CCl}$

$bh_{C\text{-}C}$=-149.777	$bh_{C\text{-}Cl}$=-5.0524	$bh_{O\text{-}C}$=65.0780	$bh_{O\text{-}Cl}$=1.6309	$bh_{S\text{-}C}$=13.2112	$bh_{S\text{-}Cl}$=21.7458
$ch_{C\text{-}C}$=-277.555	$ch_{C\text{-}Cl}$=3.0446	$ch_{O\text{-}C}$=-23.3729	$ch_{O\text{-}Cl}$=9.7281	$ch_{S\text{-}C}$=-52.3605	$ch_{S\text{-}Cl}$=76.4933
$dh_{C\text{-}C}$=300.7551	$dh_{C\text{-}Cl}$=-21.8810	$dh_{O\text{-}C}$=193.8232	$dh_{O\text{-}Cl}$=-2.7347	$ch_{S\text{-}C}$=-219.451	$dh_{S\text{-}Cl}$=-171.423

Finally, substituting all the AIPs, the number of atoms and the CIs values, the value of the DMSO-CCl group interaction parameter obtained is $a_{DMSO\text{-}CCl}$ = 287.05.

The UNIFAC parameter table for the 1,2-dichloroethane/DMSO is given in Table 5.12.

Table 5.12 Filled UNIFAC parameter table for the system 1,2-dichloroethane/DMSO in the reference UNIFAC model (original UNIFAC[71])

	CCl	DMSO
CCl	0.0	-141.77
DMSO	287.05	0.0

A comparison between experimental data and calculated values using UNIFAC-CI for the system 1,2-dichloroethane/DMSO at 95.3 kPa, is shown in Figure 5.14. The agreement between experimental data and the predicted values is good. There are small discrepancies for the bubble points at low compositions of 1,2-dichloroethane, but overall the prediction is satisfactory.

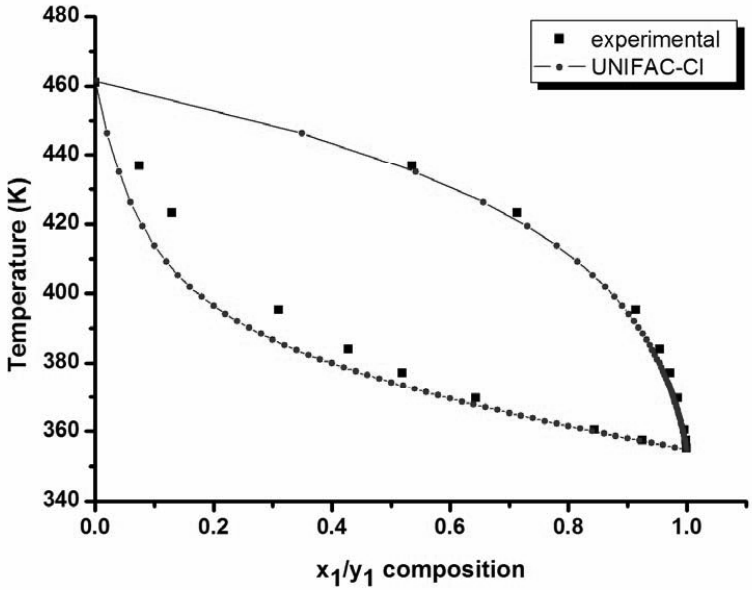


*Figure 5.14 UNIFAC-CI prediction for the system 1,2-dichloroethane (1) – DMSO (2). P = 95.3 kPa. Comparison between UNIFAC-CI (-●-) and experimental data (■). Data from Radhamma M. et al. J. Chem. Eng. Data, **53**, 374 (2008).*

5.4 Discussion: Letter to the Editor of the AIChE Journal

Addressing the criticism of the proposed UNIFAC-CI methodology, which was the topic of the paper by Mohs et al. (Mohs A., Jakob A., Gmehling J., "Analysis of a Concept for Predicting Missing Group Interaction Parameters of the UNIFAC Model Using Connectivity Indices", *AIChE J.* 2009-in press), accepted for publishing in the AIChE Journal, given here, in its entirety, is the letter prepared to follow that article. The objective of the letter is to clarify the issues raised by Mohs et al, to point out their use of the UNIFAC-CI method in a manner not recommended in our paper (González et al, AIChE J., 2007) and to provide the reader some of our results for the same systems Mohs et al did not find the UNIFAC-CI to do well.

In the paper published in the AIChE journal, 'Analysis of a concept for predicting missing group interaction parameters of the UNIFAC model using connectivity indices' by A Mohs, A Jakob, J Gmehling, the authors have examined the reliability of our method (González H E, Abildskov J, Gani R, Rousseaux P, Le Bert B. 'A method for prediction of UNIFAC group interaction parameters'. *AlChE J*.53 (2007); 1620-1633). We thank the authors for undertaking this work and appreciate their efforts to examine the reliability of our method. We do not question the accuracy of their calculations. We also agree with the authors that filling the group interaction parameters with experimental data is the best option. What we pointed out in our paper, however, is that when the experimental data is not available, our UNIFAC-CI method provides a fast and simple estimation of the parameters. The accuracy, however, like the group contribution method itself, or the COSMO-RS method, can be good in some cases but not in every case.

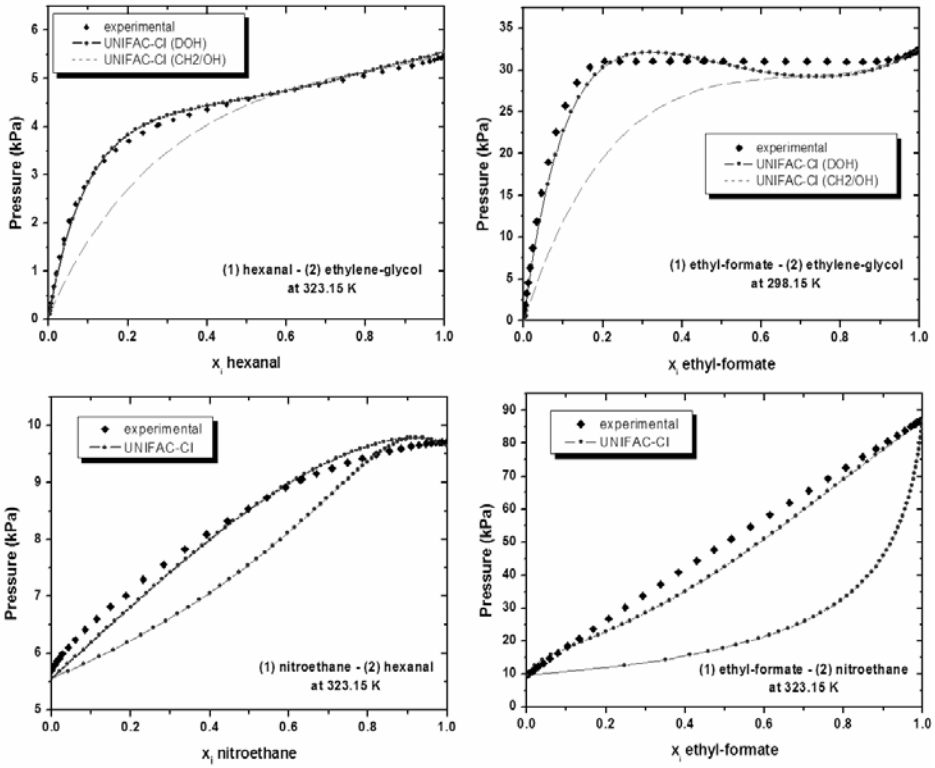
What the Mohs *et al*. paper has illustrated very well are the following two issues. A) Should the atom interaction parameters (AIPs) for functional groups be used for estimating the molecular group interaction parameters? B) Which types of new groups can be created with the AIPs for the functional groups? From the examination made by Mohs *et al*., it is clear that the AIPs for functional groups should not be used to fill-out the interactions involving the molecular groups. This is also the reason why we provided separate AIP tables involving the molecular groups, water and methanol. We acknowledge that some confusion could have been created by case study 3 in our paper where we considered the system 1,2-ethanediol-methanol (see Figure 9 in our paper). For this case study, we actually used the AIPs for the methanol molecular group and not the functional AIPs. Also from the examination made by Mohs *et al.*, it is clear that one should be careful about creating totally new groups. In our group parameter table for the reference UNIFAC model, we did not have groups 51 and above. Therefore, we did not also include any data for systems that required these groups. Also, we clearly pointed out in our paper that we excluded the groups COO and CON (on page 1624; see also the Table 2 on this page) from our correlation of the functional AIPs. It is unfortunate, therefore, that Mohs *et al.* used group 46 (CON) and group 57 (ACCN) for the evaluation of our UNIFAC-CI method. It is therefore not surprising that our published AIPs did not give acceptable results for systems involving these groups.

As Mohs *et al.* also found, our method appears to perform well for the groups included in the parameter table of the reference UNIFAC model we used to develop the corresponding AIP tables. The AIPs we provided can be used mainly to fill the empty spaces in the reference UNIFAC model parameter table. Also, new groups can be

created but these groups, in principle, should be combinations of existing groups – a point that we did not explicitly mention but implied through our case study 3. Totally new groups can be created with more confidence, if more representative data could be used in estimating the AIPs, which is also true for all group contribution based methods like UNIFAC. Again, the poor performance reported by Mohs *et al*. is not surprising. Clearly, we did not recommend (or expect) this kind of extrapolations of our published AIPs.

Since publication of our original AIPs, we have revised and enlarged our parameter tables by adding new atoms (González H E, Abildskov J, Gani R, 'Computer-aided framework for pure component properties and phase equilibria prediction for organic systems', *Fluid Phase Equilibria*, **261**, 199 (2007)). Just like Prof. Gmehling and his group, we also regularly revise our parameter tables. With our latest parameter tables (not yet published) we can represent the four systems corresponding to Figs 18a-18d of Mohs *et al.*, very well (see Figures 1a-1d below). Here, we have used AIPs regressed for the molecular group DOH as well as represented ethylene glycol with groups CH2 and OH (and used our AIPs for functional groups). For the two nitroethane systems, we also provide the calculated vapor compositions in the corresponding figures. In Table 1, we provide the corresponding group interaction parameters (GIPs) calculated from our latest AIPs.

In Figures 2a and 2b, we provide the correlation statistics for systems with C, H, O and N atoms. The systems with the highest errors relate to acetone-aniline (CH_2, ACH, CH_2CO, $ACNH_2$), dimethylamine-n-hexane (CH_2, C_2NH) and acetonitrile-ethanol (CH_2, OH, CCN). If one takes these systems and extrapolates the AIPs, the results may not be very good. The same may also be true for the systems where the reference-UNIFAC model does not represent the data very well.



Figures 1a-1d: Comparison of the UNIFAC-CI method calculations with VLE data (Mohs et al) for ethylene glycol (top two figures) and nitroethane (bottom two figures) with 1-hexanal and ethyl formate.

Table 1: The calculated GIPs estimated through our latest AIPs with our UNIFAC-CI method (used to calculate the phase diagrams in Figure 1).

GIPs	Parameter a_{mn}	Parameter a_{nm}
CH2-CHO	292.37	32.67
CH2-HCOO	615.28	102.41
CH2-CNO2	-21.81	-151.18
CH2-DOH	1154.34	69.23
CHO-CNO2	7948.11	2510.69
CHO-DOH	813.73	-512.43
HCOO-CNO2	-360.04	34425.70
HCOO-DOH	53.19	377.14

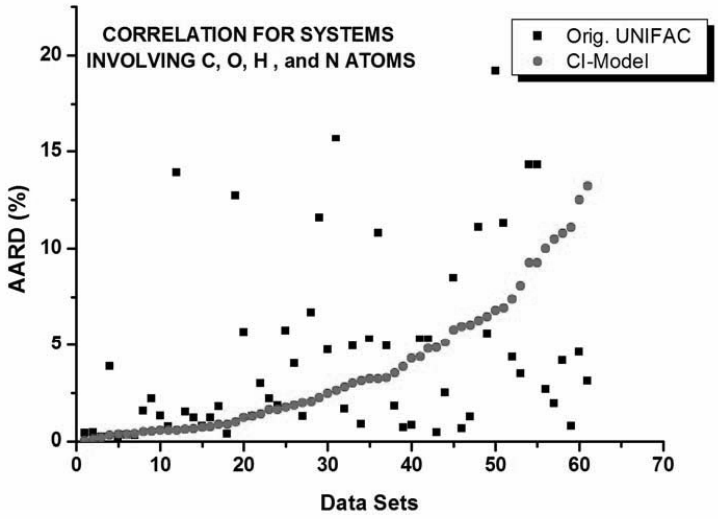


Figure 2a: Correlation for systems with C, H, O, and N atoms for the reference UNIFAC model and the UNIFAC-CI method.

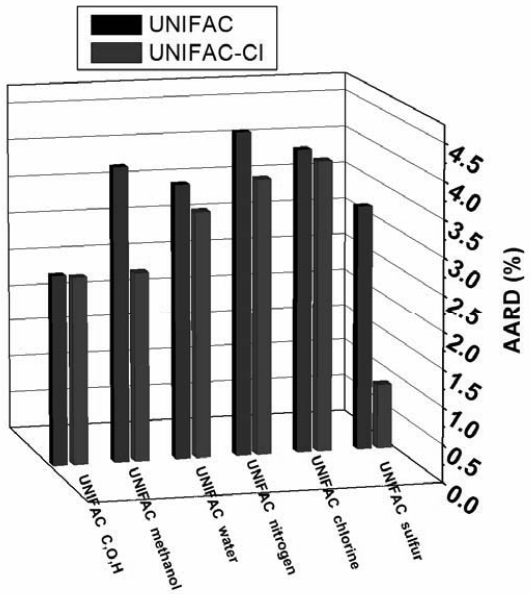


Figure 2b: Comparison of the reference UNIFAC model and the UNIFAC-CI method.

Finally, it is true that for molecular group interactions (pointed out by Mohs et al. for the water-methanol system), there can be two sets of generated parameters. That is why in our latest revision, we have forced these parameters to be the same. Also, just like a molecule can be represented differently by different groups, the same problem appears when representing groups with the atom-CI, therefore, this is not necessarily a limitation for only the atom-CI representation of groups. For this reason, we have developed rules for group as well as atom-CI representation, providing thereby, a unique representation.

Our current work is expanding the application of the UNIFAC-CI method to other reference UNIFAC models (such as Original UNIFAC (LLE) and Modified UNIFAC-Dortmund) as well as examining in more detail, the systems where the AIPs work and the systems where the AIPs do not work. We thank Mohs et al. for their interest in our UNIFAC-CI method and we hope they will find our method to perform better than they have experienced when they use our latest AIPs.

Chapter 6

Conclusions

6.1 DISCUSSION

Based on the GC^{Plus} concept, a methodology for generating UNIFAC GIPs through CIs has been established. The use of CIs for correlating mixture properties (VLE, SLE, LLE) has been highlighted through successful correlation results. The original UNIFAC-CI (VLE and LLE) and the Modified UNIFAC-CI (Dortmund) GC^{Plus} models have been formulated, analyzed and developed. In Chapter 2, the background related to the GC^{Plus} model was given, together with a functional analysis of the significance and convenience of the use of CIs to describe UNIFAC groups. In Chapter 3, both models were derived, the rules and several illustrative examples for their correct use were given.

The UNIFAC-CI models, in general, allow generation of the missing (or already existing) group interactions based on structural information of the UNIFAC groups that are described by CIs and the stoichiometric values of the groups. These quantities and the GIPs are linked by adjustable parameters (AIPs). Using the information concerning stoichiometry, CIs, and AIPs needed, then allows the following possibilities:

1) The generation of missing UNIFAC GIPs;
2) The creation of GIPs involving new groups that have not been previously calculated.

In the early stage of this work, several investigations were performed relate to the application of a successful optimization scheme for generation of the needed AIPs, until the best possible fit has been achieved. Serial and in 'parallel' approaches for regressing the AIPs were tested for a number of optimization runs. This has lead to a set of basic rules that are the basis for future studies for the estimation of the best values of AIPs.

The obtained AIPs were tested through interpolative/extrapolative predictions in Chapter 5, where their performance was compared to experimental results. Good results were obtained in the interpolative region, with some exceptions that were discussed. The extrapolation in the use of the AIPs was done through predictions of SLE for several systems and for generation of missing GIPs that were created and evaluated against experimental data were given. At present, the number of compounds, which have been tested with the UNIFAC-CI models is not large. A definite statement related to the application of these methods in terms of classes of compounds, temperature, and pressure ranges is difficult to make. The results are however, encouraging and support the idea that the UNIFAC-CI models can be used to generate missing UNIFAC GIPs when no experimental information is available.

In Chapter 5, a comprehensive discussion, regarding the limitations, the use, and the reliability of the UNIFAC-CI models has shown how this work can be improved, but at the same time described and gave a non-bias clarification of the uses and misuses of the UNIFAC-CI models.

6.2 CONTRIBUTIONS OF THIS THESIS

- The GC^{Plus} methodology has been applied successfully for three GC models (original UNIFAC-VLE and its extrapolation to SLE, original UNIFAC-LLE, and Modified UNIFAC (Dortmund)), for mixture properties;
- A computer-aided framework for the development of GC^{Plus} models for mixture properties has been proposed for several models, covering an extensive application range;
- It has been demonstrated that the CIs can be used to correlate mixture properties successfully, with many potential applications;
- A non-linear optimization problem involving an interdependency between AIPs and GIPs has been formulated and solved to find the optimal AIPs needed to generate the missing GIPs;
- A systematic and rational approach has been formulated for the interpolative/extrapolative predictions using UNIFAC group interaction parameters for original UNIFAC (VLE and LLE) and Modified UNIFAC (Dortmund);
- It has been proven that, if adequately used, the UNIFAC-CI models can be an option for filling in the missing gaps in the UNIFAC parameter tables, when experimental data are not available.

6.3 FUTURE WORK

Increasing the diversity of the experimental data system used for regressing the AIPs

Obtaining better predictions using UNIFAC models with the generated GIPs depends strongly on the diversity of the chemical classes present in the data sets used for correlation. Consider for example, a data system used for correlation where there are data sets containing classes of compounds of types A, B, C and D. If a majority of the data sets belong to classes A and C, then the predictions will be better for those data sets with chemical similarities to types A and C.

If the system of data sets used for regressing the AIPs for the groups containing nitrogen in the Original UNIFAC-CI is inspected, the following statistics can be extracted; as shown in Figure 6.1.

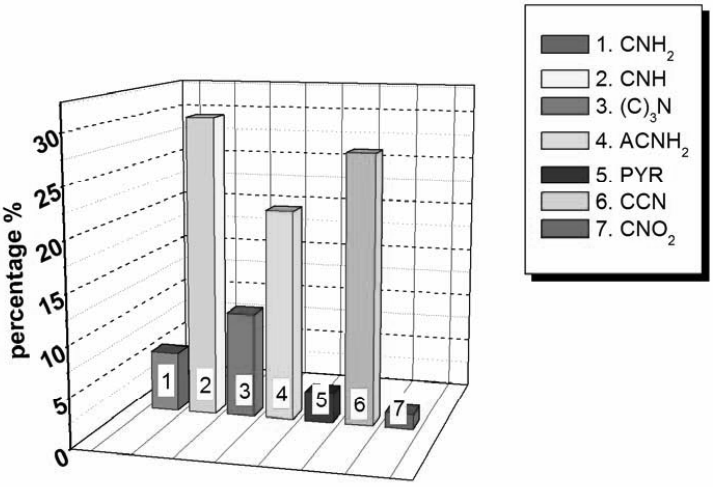


Figure 6.1 Distribution (in percentage) of the nitrogenated UNIFAC main groups involved in the data sets used for correlation of the AIPs for UNIFAC-CI (VLE).

It is clear that a majority of the data sets used contain the UNIFAC groups CNH, $ACHN_2$, and CCN while the CNO_2 group is represented in the smaller proportion. It can be said that the optimized AIPs will give better predictions for the systems with the characteristic behavior of the the UNIFAC main groups CNH and CCN than, for group, CNH_2. Experience suggests that the richer (not necessarily in number) the data system for correlation, the better the predictions that can be achieved for a broader diversity of mixtures of different chemical compounds. It should be mentioned that none of the other nitrogen-containing UNIFAC groups from the reference UNIFAC model[6], that is, ($ACNO_2$, ACRY, DMF, NMP, CON, and MORPHOLINE), were taken into account. All these reasons support the need for making a selective collection of reliable experimental data that should represent a broader set of UNIFAC groups for the specific class of chemicals under study.

New types of experimental data are needed

In the case of Original UNIFAC-CI (LLE) and Modified UNIFAC-CI (Dortmund), it is necessary to consider different types of data in order to 'enrich' the optimized AIPs. In order to make predictions of LLE, SLE, and h^{EX} with Modified UNIFAC-CI (Dortmund), it is necessary to use the same kind of experimental data as used for the development of the reference model[73]. This means that LLE, h^{EX}, SLE, γ^{∞}, $c_P{}^E$ should be collected and prepared to be implemented in the UNIFAC-CI parameter optimization software.

New parameter optimization schemes

The use of new parameter optimization routines should be considered, and in spite of the problem size when regressing UNIFAC-related parameters, global optimization schemes should be considered at least for some stages of the entire parameter optimization process. Investigations should be carried out in terms of computing time and numerical limits. The possibility of matching the AIPs with the original UNIFAC parameters instead of experimental data should also be considered. This approach would need to give good results that could be taken afterwards as reliable initial estimates for the parameter optimization process using experimental data. All the possible difficulties and changes needed in the optimization code should be evaluated.

Statistical Analysis

A statistical analysis involving hundreds and possibly thousands of data sets for validation against experimental data would be valuable in order to define trends in the predictions using a defined set of AIPs. It might also provide evidence of the changes needed in order to improve the predictive power of the UNIFAC-CI models. Investigations comparing the predictions using full and partially CI-generated GIPs are necessary to make conclusions and recommendations for the user.

Numerical Aspects

There are several quantities, numbers, schemes, and representations involved in the estimation of the AIPs that influence the quality of the parameter optimization. The best possible combinations should be explored in a systematic way, and the needed changes should be considered. Some of the recommendations for future use are the following: 1) the use of other CI values within the Kier-Hall theory (or even other theories) offer challenging options to develop and improve the methodology; 2) the regularization schemes to accelerate the convergence of the optimization sometimes generates solutions that are far from being the best ones, this effect is closely related to the initial guesses used. This functionality should be systematically analyzed; 3) When generating the optimal values for the AIPs, sometimes very high positive and negative values are generated for the GIPs. Such high values are supported for the different kinds of combinations with the other generated GIPs that create an effect of ‘compensation’ for such big values. While for correlation effects these do not pose big difficulties, when generating new GIPs for predictions, such a high values (specially the negative ones) can lead to very poor predictions. It should be possible to ‘bound’ the generated GIPs to rational limits when using AIPs in order to avoid the above mentioned difficulties.

New Measurements

Gaps in the UNIFAC parameter tables exist mainly for the lack of experimental data. Systematic measurements of phase equilibria data can lead to improved correlations for specific families of chemical compounds that can then be safely extrapolated. At the same time, there are certain pairs of UNIFAC interactions for which experimental data is not available for validation purposes, key systems can provide valuable information that can be used for UNIFAC groups (and its interactions) that behave in a similar chemical way.

List of Symbols

a_{mn}	UNIFAC group interaction parameter of groups m-n
A_{wk}	van der Waals group surface area
$A_{m,i}$	Number of atoms of type *i* occurring within a missing group/fragment s*
$a_{m,i}$	Contribution of atom *i* within a missing group/fragment s*
a_i	Activity
$\left(A_{mn}^{X-Y}\right)_n$	Coefficient involving CIs and stoichiometric information, (Eqs. 3.3-3.6)
b	Adjustable parameter in Equation (2.22)
b_{X-Y}	Atom interaction parameter of level 1 for the atoms *X* and *Y*
$\overline{b_{X-Y}}$	Atom interaction parameter (mirror) of level 1 for the atoms *X* and *Y*
b_{mn}	UNIFAC group interaction parameter between the m-n groups
b_{X-Y}^{1}	Atom interaction parameter of level 1 for the atoms *X-Y*, corresponding to b_{mn}
$\overline{b_{X-Y}^{1}}$	Atom interaction parameter of level 1 (mirror) for the atoms *X-Y*, corresponding to b_{mn}
b_{X-Y}^{2}	Atom interaction parameter of level 1 for the atoms *X-Y*, corresponding to c_{mn}
$\overline{b_{X-Y}^{2}}$	Atom interaction parameter of level 1 (mirror) for the atoms *X-Y*, corresponding to c_{mn}
c_{mn}	UNIFAC group interaction parameter between the m-n groups
bh_{X-Y}	Atom interaction parameter (mirror), Appendix B
C	Total number of compounds in Equation (1.14)
C_i	Number of occurrences of a group in Equation (2.1)
Cp	Heat capacity
C_i	First order group of type *i* in Equation (2.3)
C_k	$\left(\delta_i^v \delta_j^v\right)_k^{-1/2}$ for each bond *k* in a molecule/group
c	Adjustable parameter in Equation (2.22)
c_{X-Y}	Atom interaction parameter of level 2 for the atoms *X* and *Y*
$\overline{c_{X-Y}}$	Atom interaction parameter (mirror) of level 2 for the atoms *X* and *Y*
c_{mn}	UNIFAC group interaction parameter between the m-n groups
c_{X-Y}^{1}	Atom interaction parameter of level 2 for the atoms *X-Y*, corresponding to b_{mn}
$\overline{c_{X-Y}^{1}}$	Atom interaction parameter of level 2 (mirror) for the atoms *X-Y*, corresponding to b_{mn}
c_{X-Y}^{2}	Atom interaction parameter of level 2 for the atoms *X-Y*, corresponding to c_{mn}
$\overline{c_{X-Y}^{2}}$	Atom interaction parameter of level 2 (mirror) for the atoms *X-Y*, corresponding to c_{mn}
Cp^E	Excess heat capacity
D_j	Second order group of type *i* in Equation (2.3)
D_m	$\left(\delta_i^v \delta_j^v \delta_k^v\right)_m^{-1/2}$ for each three-atom-vertex bond *m* in a molecule/group
d	Adjustable parameter in Equation (2.23)
d_{X-Y}	Atom interaction parameter of level 3 for the atoms *X* and *Y*
$\overline{d_{X-Y}}$	Atom interaction parameter (mirror) of level 3 for the atoms *X* and *Y*

d^1_{X-Y}	Atom interaction parameter of level 3 for the atoms *X*-*Y*, corresponding to b_{mn}
$\overline{d^1_{X-Y}}$	Atom interaction parameter of level 3 (mirror) for the atoms *X*-*Y*, corresponding to b_{mn}
d^2_{X-Y}	Atom interaction parameter of level 3 for the atoms *X*-*Y*, corresponding to c_{mn}
$\overline{d^2_{X-Y}}$	Atom interaction parameter of level 3 (mirror) for the atoms *X*-*Y*, corresponding to c_{mn}
E_k	Third order group of type *i* in Equation (2.3)
e_{X-Y}	Atom interaction parameter of level 4 for the atoms *X* and *Y*
$\overline{e_{X-Y}}$	Atom interaction parameter (mirror) of level 4 for the atoms *X* and *Y*
e^1_{X-Y}	Atom interaction parameter of level 4 for the atoms *X*-*Y*, corresponding to b_{mn}
$\overline{e^1_{X-Y}}$	Atom interaction parameter of level 4 (mirror) for the atoms *X*-*Y*, corresponding to b_{mn}
e^2_{X-Y}	Atom interaction parameter of level 4 for the atoms *X*-*Y*, corresponding to c_{mn}
$\overline{e^2_{X-Y}}$	Atom interaction parameter of level 4 (mirror) for the atoms *X*-*Y*, corresponding to c_{mn}
f	Fugacity
f_i°	Fugacity of species type *i* in standard state in Equation (1.25)
f(x)	Property function in Equation (2.2)
G	Gibbs Free Energy
h_i	number of bonded hydrogen on atom *i*
h^E	Molar excess enthalpy
M_i	Group 'M' in Equation (2.1)
M_j	Number of occurrences of second order in Equation (2.3)
m	Group of type *m*
n_i	Composition
N_i	Group contribution of type *i* in Equation (2.2)
N_i	Number of occurrences of first order in Equation (2.3)
n_i	lone-pair electrons on atom *i*
n	Refractive index
n	Group of type *n*
N_H	Number of hydrogen atoms attached to each vertex-atom
n_x	Number of atoms of type X present in an UNIFAC group
O_k	Number of occurrences of third order in Equation (2.3)
P	Pressure
Pc	Critical Pressure
p_i	Π orbital electrons on atom *i*
$P_{i\text{-}calc}$	Calculated pressure for the data point *i* in Equation (4.1)
$P_{i\text{-}exp}$	Calculated pressure for the experimental data point *i* in Equation (4.1)
U	Internal Energy
S	Entropy
W	Work
V	Volume
$\underline{V}$	Molar volume
R	Gas constant
P_n	Specific property 'n' in Equation (2.1)
Q_k	Surface area parameter, for group *k*
q_i	Molecular surface area parameter, for component *i*
R_k	Volume parameter, for group *k*
r_i	Molecular volume parameter, for component *i*
Tc	Critical Temperature

T	Temperature
Tg	Glass transition temperature
Vc	Critical Volume
V_{wk}	van der Waals group volume area
V_g	Glassy amorphous volume
V_r	Rubbery amorphous volume
V_a	Amorphous volume
V_c	Crystalline volume
w_{reg}	Regularization factor in Equation (4.3)
x_i	Mole fraction of component I (liquid)
X	Atom of type X
x_i	Liquid phase composition
x_i^{calc}	Calculated liquid phase composition
x_i^{exp}	Experimental liquid phase composition
Y_m	Pure component property corresponding a missing group/fragment *m*
Y*	Added contribution within a missing group/fragment s* for a property *Y*
Y	Atom of type Y
y_i	Vapor phase composition
z	lattice coordination number
Z^v	See Equation (2.20)
Z^V	Number of valence electrons in the atom

Greek letters

α	Phase alfa in Equation (1.20)
β	Phase beta in Equation (1.20)
β^k	Product of δ^v for atom-vertices i-j, $\delta_i^v \delta_j^v$
γ^k	Product of δ^v for atom-vertices i-j-k, $\delta_i^v \delta_j^v \delta_k^v$
γ^∞	Activity coefficient at infinite dilution
γ	Activity coefficient
μ	Internal Energy
Π	Total number of phases in Equation (1.14)
μ°	Integration constant in Equation (1.18)
φ	Fugacity coefficient
ΔH_f	Enthalpy of formation
ΔG_f	Gibss free energy
Φ_i	Volume fraction for component I in mixture
θi	surface are fraction, for i in mixture
θji	surface area fraction, local for j around i
ν_{ki}	number of groups k present in molecule i
Γk	Residual group activity coefficient for group k
Γ_k^i	Residual group activity coefficient, in pure component i
Θ	Area fraction
${}^n\chi^v$	Valence connectivity indices
δ	number of bounded neighbors
δ_i^v	number of valence electrons of atom vertex i
σ_i	σ orbital electrons on atom *i*

δ	Solubility parameter
I	Phase I in Equation (4.4)
II	Phase II in Equation (4.4)

Subscripts

c	Critical property
C	Combinatorial in Equation (2.5)
calc	Calculated value
exp	Experimental value
i	Species type in Equation (1.12)
i	Compound i
k	Group k
k	Each of the phase in Equation (1.12)
mn	Group interaction between groups m and n
R	Residual in Equation (2.5)
X-Y	Interaction between atom *X* and atom *Y*

Superscripts

°	Standard state
∞	at infinite dilution
*	Missing group/fragment
E	Excess property

Abbreviations

ASOG	Analytic Solution of Groups model
AIP	Atom interaction parameters
AARD	Average Absolute Relative Deviation
CI	Connectivity Index
CPA	Cubic Plus Association
CAS	Chemical Abstracts Service
COSMO-RS	Conductor-like Screening Model for Real Solvents
CODESSA	Comprehensive Descriptors for Structural and Statistical Analysis
DFT	Density Functional Theory
EoS	Equation of State
FH	Flory-Huggins
GIP	Group Interaction Parameter
GC^{Plus}	Group Contribution Plus method
GC-EoS	Group Contribution – equation of state
IG	Initial guesses
LLE	Liquid-Liquid Equilibrium
MOSCED	MOdified Separation of Cohesive Energy Density
NRTL	Non Random Two Liquid model
NG	Total number of groups in Equation (2.1)
ND	Number of data points

N	Number of experimental data points in Equation (4.1)
OF	Objective function
POY	Poynting factor correction
PSRK	Predictive Soave-Redlich-Kwong
QSAR	Quantitative Structure Activity Relationship
QSPR	Quantitative Structure Property Relationship
SLE	Solid-Liquid Equilibrium
SARE	Sum of Absolute Relative Error
SRE	Sum of Relative Error
UNIQUAC	Universal Quasi-Chemical model
VLE	Vapor-Liquid Equlibrium
WILSON	WILSON model

REFERENCES

1. Abildskov J., *Extended Group Contribution Model for Polyfunctional Phase Equilibria*, PhD Thesis, Institut for Kemiteknik, Technical University of Denmark, Lyngby, Denmark (1998).

2. Pierroti G. J., Deal C. H., and Derr E. L., "Activity Coefficients and Molecular Structure*"*, *Ind. Eng. Chem.*, **51**, 95 (1959).

3. Thomas E. R., and Eckert C. A., "Prediction of Limiting Activity Coefficients by a Modified Separation of Cohesive Energy Density Model and UNIFAC", *Ind. Eng. Chem. Proc. Des. Dev.*, **23**, 194 (1984).

4. Guggenheim E., *Mixtures*, Clarendon Press, Oxford (1952).

5. Derr E. L., and Deal C. H., "Analytical Solutions of Groups. Correlation of Activity Coefficients through Structural Group Parameters", *I.Chem.E. SYMPOSIUM SERIES*, ***No. 32***, 340 (1969).

6. Fredenslund Aa., Jones R. L., and Prausnitz J. M., "Group contribution estimation of activity coefficients in nonideal liquid mixtures", *AIChE Journal*, **21**, 1086 (1975).

7. Gani R., Harper P., and Hostrup M., "Automatic Creation of Missing Groups through Connectivity Indices for Pure-Compound Property Prediction", *Ind. Eng. Chem. Res.*, **44**, 7262 (2005).

8. Satyanarayana K. C., Gani R., and Abildskov J., "Polymer Property Modeling using Grid Technology for Design of Structured Products", *Fluid Phase Equilibria*, **261**, 58 (2007).

9. Poling B. E., Prausnitz J. M., and O´Connell J. P., *The Properties of Gases and Liquids*, McGraw-Hill, New York (2001).

10. Chen C.-C., and Mathias P.M., "Applied thermodynamics for process modeling", *AIChE Journal,* **48**, 194 (2002).

11. IUPAC Compendium of Chemical Terminology: http://goldbook.iupac.org/.

12. CAS registry (http://www.cas.org/expertise/cascontent/registry/index.html).

13. Lydersen A. L., *Estimation of critical properties of organic compounds*, College Engineering University Wisconsin, *Engineering Experimental Station Report 3*, Madison, Wisconsin (1955).

14. Ambrose D., *Correlation and estimation of vapor–liquid critical properties. I.Critical temperatures of organic compounds*, National Physical Laboratory, Teddington, UK, *NPL Report Chem.*, 92, (1978).

15. Klincewicz K. M., and Reid R. C., "Estimation of Critical Properties with Group Contribution Methods", *AIChE Journal,* **30**, 137 (1984).

16. Joback K. G., and Reid R. C., "Estimation of Pure-Component Properties from Group-Contributions", *Chem. Eng. Comm.*, **57**, 233 (1987).

17. Constantinou L., and Gani R., "New Group Contribution Method for Estimating Properties of Pure Components", *AIChE Journal,* **40,** 1697 (1994).

18. Marrero J., and Gani R., "Group Contribution Based Estimation of Pure Component Properties", *Fluid Phase Equilibria*, **183–184**, 183 (2001).

19. Larsen B. L., Rasmussen P., and Fredenslund Aa., "A Modified UNIFAC Group-Contribution Model for the Prediction of Phase Equilibria and Heats of Mixing", *Ind. Eng. Chem. Res.*, **26**, 2274 (1987).

20. Weidlich U., and Gmehling J., "A Modified UNIFAC Model. 1- Prediction of VLE, h^E, and γ^{∞}", *Ind. Eng. Chem. Res.*, **26**, 1372 (1987).

21. Kehiaian H. V., "Group contribution methods for liquid mixtures: A critical review", *Fluid Phase Equilibria*, **13**, 243 (1983).

22. Fredenslund Aa., and Rasmussen P., "From UNIFAC to SUPERFAC - and back?", *Fluid Phase Equilibria*, **24**, 115 (1985).

23. Abildskov J., Constantinou L., and Gani R., "Towards the development of a second-order approximation in activity coefficient models based on group contribution", *Fluid Phase Equilibria*, **118**, 1 (1996).

24. Dahl S., and Michelsen M. L., "High-pressure Vapor-liquid Equilibrium with a UNIFAC-based Equation of State", *AIChE Journal*, **36**, 1829 (1990).

25. Dahl S., Fredenslund Aa., and Rasmussen P., "The MHV2 Model: a UNIFAC-based Equation of State Model for Prediction of Gas Solubility and Vapor-liquid Equilibria at Low and High Pressures", *Ind. Eng. Chem. Res.*, **30**, 1936 (1991).

26. Holderbaum T., and Gmehling J., "PSRK: A Group Contribution Equation of State Based on UNIFAC", *Fluid Phase Equilibria*, **70**, 251 (1991).

27. Skjold-Jørgensen S., "Gas solubility calculations. II. Application of a new group-contribution equation of state", *Fluid Phase Equilibria*, **16**, 317 (1984).

28. Tamouza S., Passarelo J. P., Tobaly P., and de Hemptinne J. C., "Group contribution method with SAFT EoS applied to vapor liquid equilibria of various hydrocarbon series", *Fluid Phase Equilibria,* **222**, 67 (2004).

29. Tamouza S., Passarelo J. P., Tobaly P., and de Hemptinne J. C., "Application to binary mixtures of a group contribution SAFT EoS (GC-SAFT)", *Fluid Phase Equilibria*, **228**, 409 (2005).

30. Thi T. X. N., Tamouza S., Tobaly P., Passarelo J. P., and de Hemptinne, J. C., "Application of group contribution SAFT equation of state (GC-SAFT) to model phase behaviour of light and heavy esters", *Fluid Phase Equilibria*, **238**, 254 (2005).

31. Le Thi C., Tamouza S., Passarello J. P., Tobaly P., and de Hemptinne J. C., "Modeling Phase Equilibrium of H_2 + *n*-Alkane and CO_2 + *n*-Alkane Binary Mixtures Using a Group Contribution Statistical Association Fluid Theory Equation of State (GC-SAFT-EoS) with a k_{ij} Group Contribution Method", *Ind. Eng. Chem. Res.*, **45**, 6803 (2006).

32. Lymperiadis A., Adjiman C. S., Galindo A., and Jackson G., "A Group Contribution Method For Associating Chain Molecules Based On The Statistical Associating Fluid Theory SAFT-γ", *J. Chem. Phys.*, **127**, 234903 (2007).

33. Wolbach J. P., and Sandler S. I., "Using Molecular Orbital Calculations To Describe the Phase Behavior of Hydrogen-Bonding Fluids", *Ind. Eng. Chem. Res.*, **36**, 4041 (1997).

34. Klamt A., and Eckert F., "COSMO-RS: a novel and efficient method for the *a priori* prediction of thermophysical data of liquids", *Fluid Phase Equilibria*, **172**, 43 (2000).

35. Arlt W., Spuhl O., and Klamt A., Challenges in Thermodynamics, *Chem. Eng. and Proc.*, **43**, 221 (2004).

36. Ravindranath D., Neely B. J., Robinson Jr. R. L., and Gasem K. A. M., "QSPR Generalization Of Activity Coefficient Models For Predicting Vapor–Liquid Equilibrium Behavior", *Fluid Phase Equilibria*, **257**, 53 (2007).

37. CODESSA-PRO, QSPR/QSAR Software, (http://www.codessa-pro.com/).

38. Bunz A. P., Braun B., and Janowsky R., "Quantitative structure–property relationships and neural networks: correlation and prediction of physical properties of pure components and mixtures from molecular structure", *Fluid Phase Equilibria*, **158–160**, 367 (1999).

39. Kasturirangan A., *Cubic equation-of-state modeling of asymmetric binary mixtures*, MSc Thesis, Oklahoma State University. Oklahoma (2004).

40. Benson S. W., and Buss J., "Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties", *The Journal of Chemical Physics*, **9**, 3 (1958).

41. Franklin J. L., "Prediction of Heat and Free Energies of Organic Compounds", *Industrial and Engineering Chemistry*, **41**, 5 (1949).

42. Verma K. K., and Doraiswamy, L. K., "Estimation of Heats of Formation of Organic Compounds", *Ind. Eng. Chem. Fundam.* **4**, 389 (1965).

43. van Krevelen D. W., and Chermin H. A. G., "Estimation of the Free Enthalpy (Gibss Free Energy) of Formation of Organic Compounds from Group Contribution", *Chem. Eng. Sci.*, **1**, 68 (1951).

44. van Krevelen D. W., and Chermin H. A. G., "Erratum: Estimation of the Free Enthalpy (Gibss Free Energy) of Formation of Organic Compounds from Group Contribution", *Chem. Eng. Sci.*, **1** (1952).

45. Thomas L. H., *J. Chem. Soc.* 820 (1946).

46. Orrick C., and Erbar, *Estimation of viscosity for organic liquids*, Proposition Report, Oklahoma State University, Stillwater, Oklahoma (1973).

47. van Velzen D., Cardozo R. L., and Langemkamp H., "A Liquid Viscosity-Temperature-Chemical Constitution Relation for Organic Compounds", *Ind. Eng. Chem. Fundam*., **11** (1972).

48. Morris P. S, MSc Thesis, Polytechnic Institute of Brooklyn, Brooklyn, N.Y. (1964).

49. Langmuir I., *The Distribution and Orientation of Molecules*, Third Colloid Symposium Monograph, The Chemical Catalog Company Inc., New York (1925).

50. Redlich O., Derr E. L., and Pierotti G., "Group Interaction I. A Model for Interactions in Solution", *J. Am. Chem. Soc.*, **81**, 2283 (1959).

51. Derr E. L., and Papadopoulos M., "Group Interaction II. A test of the Group Model on Binary Solutions of Hydrocarbons", *J. Am. Chem. Soc.*, **81**, 2285 (1959).

52. Ronc M., and Ratcliff G. A., "Prediction of Excess Free Energies of Liquid Mixtures by an Analytic Group Solution Model", *Can. J. Chem. Eng.*, **49**, 875 (1975).

53. Kojima K., and Tochigi K., *Prediction of Vapor-Liquid Equilibria by the ASOG Method*, Kodansha-Elsevier, Tokyo (1979).

54. Abrams D. S., and Prausnitz J. M., "Statistical Thermodynamics of Liquid Mixtures. A New Expression for the Excess Gibbs Energy of Partly and Completely Miscible Systems", *AIChE Journal*, **21**, 116 (1975).

55. Fredenslund Aa., Gmehling J., and Rasmussen P., *Vapor-Liquid Equilibria Using UNIFAC,* Elsevier, Amsterdam (1977).

56. Bondi A., *Physical Properties of Molecular Crystals, Liquids and Glasses*, John Wiley & Sons, New York (1968).

57. Todeschini R., and Consonni V., *Handbook of Molecular Descriptors*, Wiley-VCH (2000).

58. Randić M., "On the characterization of molecular branching", *J. Am. Chem. Soc.*, **97**, 6609 (1975).

59. Kier L. B. and Hall L. H., "Molecular Connectivity. VII. Specific Treatment of Heteroatoms", *Journal of Pharmaceutical Sciences*, **65**, 1806 (1976).

60. Murray W. J., Hall L. H., and Kier L. B., *Journal of Pharmaceutical Sciences*, **64**, 1978 (1975).

61. Kier L. B., Murray J. W., Randić M., and Hall L. H., "Molecular Connectivity. V. Connectivity Series Applied to Density", *Journal of Pharmaceutical Sciences*, **65**, 1226 (1976).

62. Randić M., "Connectivity Index 25 Years After", *J. Mol. Graphics Modell*., **20**, 19 (2001).

63. Kier L. B., and Hall L. H., "Derivation and Significance of Valence Molecular Connectivity", *Journal of Pharmaceutical Sciences*, **70**, 583 (1981).

64. Sanderson R. T., "Electronegativities in Inorganic Chemistry II", *J. Chem. Educ.*, **31**, (1954).

65. Sanderson R. T., *Chemical Bonds and Bond Energy*, Academic, New York, N.Y. (1976).

66. Sanderson R. T., "Principles of Electronegativity. II. Applications", *J. Chem. Educ.*, **65**, 227 (1988).

67. Conte E., Martinho A., Matos H., and Gani R., "Combined Group Contribution and Atom Connectivity Index based Methods for Estimation of Surface Tension and Viscosity", *Ind. Eng. Chem. Res.*, **47**, 5751 (2008).

68. Allinger N., "Conformational Analysis. 130. MM2. A Hydrocarbon Force Field utilizing V1 and V2 Torsional Terms", *J. Am. Chem. Soc.*, **111**, 8551 (1977).

69. González H. E., Abildskov J., and Gani R., "Computer-aided Framework for Pure Component Properties and Phase Equilibria Prediction for Organic Systems", *Fluid Phase Equilibria*, **261**, 199 (2007).

70. SciFinder® Scholar 2006, American Chemical Society.

71. Hansen K. H., Rasmussen P., Fredenslund Aa., Schiller M., and Gmehling J., "Vapor-Liquid Equilibria by UNIFAC Group Contribution. 5. Revision and Extension", *Ind. Eng. Chem. Res.,* **30**, 2355 (1991).

72. Magnusson T., *Prediction of Liquid-Liquid Equilibria using UNIFAC*, PhD Thesis, Instituttet for Kemiteknik, Danmarks Tekniske Højskole, Lygnby Denmark (1980).

73. Gmehling J., Li J., and Schiller M., "A Modified UNIFAC Model. 2. Present Parameter Matrix and Results for Different Thermodynamic Properties", *Ind. Eng. Chem. Res.,* **32**, 178 (1993).

74. Van Ness H. C., "Thermodynamics in the treatment of (vapor + liquid) equilibria", *J. Chem. Thermodynamics*, **27**, 113 (1995).

75. Kier L. B., and Hall H. L., *Molecular Connectivity in Structure Activity Analysis*, John Wiley & Sons, Inc., New York (1986).

76. Gmehling J., Anderson T. F., and Prausnitz J. M. "Solid-Liquid Equilibria Using UNIFAC", *Ind. Eng. Chem. Fundam.,* **17**, 269 (1978).

77. Skjold-Jørgensen S., Kolbe B., Gmehling J., and Rasmussen P., *Ind. Eng. Chem. Proc. Des. Dev.*, **18**, 714 (1979).

78. Espada Juan J., Coto B. and Peña José L. "Liquid–liquid equilibrium in the systems furfural + light lubricating oils using UNIFAC". *Fluid Phase Equilibria*, **259**, 2, 201, (2007).

79. Pratas de Melo Maria J., Dias Ana M.A., Blesic Marijana, Rebelo Luís P.N., Vega Lourdes F., Coutinho João A.P., and Marrucho Isabel M. "Liquid–liquid equilibrium of (perfluoroalkane + alkane) binary mixtures' *Fluid Phase Equilibria*, **242**, 210 (2006).

80. Voutsas E., Louli V., Boukouvalas C., Magoulas K., and Tassios D. 'Thermodynamic property calculations with the universal mixing rule for EoS/G^E models: Results with the Peng–Robinson EoS and a UNIFAC model' *Fluid Phase Equilibria*, **241**, 216 (2006).

81. Hayden J. G., and O'Connell J. P., "A generalized method for predicting second virial coefficients", *Ind. & Eng. Chem. Proc. Des. & Develop.,* **14**, 209 (1975).

82. Marcilla A., del Mar Olaya M., Dolores Serrano M., "Comment on Liquid-Liquid Equilibrium Data Regression", *J. Chem. Eng. Data***, 52**, 2538 (2007).
32, 178-193.

83. Nielsen T.L., Abildskov J., Harper P.M., Papaeconomou I., Gani R., "The CAPEC Database". *J. Chem. Eng. Data*., 46, 1041 (2001).

84. DECHEMA Chemistry Data Series, Volume V (Parts 1 and 4), *Liquid-Liquid Equilibrium Data Collection*, ISBN 3-921567-17-3 and ISBN 3-921567-73-4

85. Kratch C., Geyer H., Ulbig P., and Schulz S., "Parameter Optimization of Group Contribution Methods in High Dimensional Solution Spaces", *Proceedings of the 6th international conference on Computational Intelligence, Theory and Applications: Fuzzy Days*, 647, Springer-Verlag, London UK, (1999).

86. Nelder J. A., Mead R. D., "A Simplex Method for Function Minimization", *Comput. J.* **7**, 308 (1965).

87. Levenberg K. A "Method for the solution of certain problems in least squares". *Quarterly App Math*., **2**, 164 (1944).

88. Marquardt D. "An algorithm for least-squares estimation of nonlinear parameters". *SIAM J of App Math*. **11**, 431 (1963).

89. González H E, Abildskov J, Gani R, Rousseaux P, and Le Bert B. "A method for prediction of UNIFAC group interaction parameters", *AIChE J*. **53**, 1620 (2007).
90. Proceedings of the 10th CAPEC annual meeting, Crown Of Scandinavia Copenhagen-Oslo, June 4-6, 2007.

91. Proceedings of the 11th CAPEC annual meeting, Kgs. Lyngby, Denmark, June 9-11 2008.

92. Radhamma M., Hsieh Cheng-Ting, Venkatesu P., Prabhakara Rao M. V., Lee Ming-Jer, and Lin Ho-mu "Isobaric Vapor–Liquid Equilibrium for Dimethylsulfoxide with Chloroethanes and Chloroethenes", *J. Chem. Eng. Data*, **53**, 374 (2008).

93. McDonald C. M., and Floudas C. A., "Global optimization and analysis for the Gibbs free energy function using the UNIFAC, Wilson, and ASOG equations", Industrial & engineering chemistry research, **34**, 1674 (1995).

94. Kemeny S., Manczinger J., Skjold-Jørgensen S., "Reductions of Thermodynamic Data by Means of the Multiresponse Maximum Likelihood Principle". *AIChE J.*, **28**, 20 (1982).

95. Fletcher R., *A modified Marquardt Subroutine for non-linear Least Squares*. Harwell Report, AERE R.6799, (1971).

LIST OF FIGURES

LIST OF TABLES

APPENDIX A

TABLES OF STOICHIOMETRIC AND CONNECTIVITY INDICES VALUES FOR THE UNIFAC GROUPS

Table A.1. Stoichiometric and CIs values for the Original UNIFAC-CI groups

Index	Group	nC	nO	nN	nCl	nS	χ^0	χ^1	χ^2
1	CH2	1	0	0	0	0	0.7071068	0.0000000	0.0000000
2	C=C	2	0	0	0	0	1.0000000	0.2500000	0.0000000
3	ACH	1	0	0	0	0	0.5773503	0.6666667	0.1924501
4	ACCH2	2	0	0	0	0	1.2071068	0.3535534	0.0000000
5	OH	0	1	0	0	0	0.4472136	0.0000000	0.0000000
6	CH3OH	1	1	0	0	0	1.4472136	0.4472136	0.0000000
7	H2O	0	1	0	0	0	0.5000000	0.0000000	0.0000000
8	ACOH	1	1	0	0	0	0.9472136	0.2236068	0.0000000
9	CH2CO	2	1	0	0	0	1.6153551	0.5576775	0.1443376
10	CHOa	1	1	0	0	0	0.9855986	0.2357023	0.0000000
11	CCOO	2	2	0	0	0	1.8164966	0.6582483	0.2874575
12	HCOO	1	2	0	0	0	1.3938469	0.4714045	0.0962250
13	CH2O	1	1	0	0	0	1.1153551	0.2886751	0.0000000
14	CNH2	1	0	1	0	0	1.5773503	0.5773503	0.0000000
15	CNH	1	0	1	0	0	1.5000000	0.5000000	0.0000000
16	(C)3N	1	0	1	0	0	1.4472136	0.4472136	0.0000000
17	ACNH2	1	0	1	0	0	1.0773503	0.2886751	0.0000000
18	PYRIDINE	5	0	1	0	0	1.9828944	0.6552630	0.2184210
19	CCN	2	0	1	0	0	1.4472136	0.4736068	0.1118034
20	COOH	1	2	0	0	0	1.3554619	0.4277309	0.0912871
21	CCl	1	0	0	1	0	2.1338934	1.1338934	0.0000000
22	CCl2	1	0	0	2	0	2.9748936	2.0874980	0.9091373
23	CCl3	1	0	0	3	0	3.9790305	1.9639610	2.2269225
24	CCl4	1	0	0	4	0	5.0355737	2.2677868	3.8571429
25	aCCl	1	0	0	1	0	1.6338934	0.5669467	0.0000000
26	CNO2	1	2	1	0	0	2.2247449	0.7415816	0.4013747
27	aCNO2	1	2	1	0	0	1.7247449	0.5374575	0.2347080
28	CS2	1	0	0	0	2	1.3164966	0.4082483	0.0833333
29	CH3SH	1	0	0	0	1	1.4472136	0.4472136	0.0000000
30	FURFURAL	5	2	0	0	0	3.4676494	1.7490568	0.9434742
31	DOH	2	2	0	0	0	2.3086408	1.1324555	0.4472136
32	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000
33	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000
34	C=-C	2	0	0	0	0	1.0000000	0.2500000	0.0000000
35	DMSO	2	1	0	0	1	2.8164966	0.9831632	0.7415816
36	ACRY	3	0	1	0	0	2.2316706	0.9205302	0.3332236
37	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000
38	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000
39	DMF	3	1	1	0	0	3.4328122	1.3883283	1.0690206
40	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000
41	COO	1	2	0	0	0	1.5245639	0.5468740	0.1290994
42	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000

43	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000
44	NMP	5	1	1	0	0	2.7997338	0.7546683	0.3108966
45	NA	0	0	0	0	0	0.0000000	0.0000000	0.0000000
46	CON	1	1	1	0	0	1.4328122	0.4939012	0.1054093
47	OCCOH	2	2	0	0	1	2.2696754	1.1049029	0.4277309
48	CH2S	1	0	0	0	1	1.4082483	0.4082483	0.0000000
49	MORPHOLINE	4	1	1	0	0	1.6820506	0.4711324	0.1318375
50	THIOPHENE	4	0	0	0	1	1.5176487	0.4572246	0.1368804

Table A.2. Stoichiometric and CIs values for the Original UNIFAC- LLE groups

Index	Group	nC	nO	nN	X0	X1	X2
1	CH2	1	0	0	0.7071068	0.0000000	0.0000000
2	C=C	2	0	0	1.0000000	0.2500000	0.0000000
3	ACH	1	0	0	0.5773503	0.6666667	0.1924501
4	ACCH2	2	0	0	1.2071068	0.3535534	0.0000000
5	OH	0	1	0	0.4472136	0.0000000	0.0000000
6	P1	3	1	0	2.8614272	1.5233345	0.7236068
7	P2	3	1	0	3.0245639	1.4128994	1.0937480
8	H2O	0	1	0	0.5000000	0.0000000	0.0000000
9	ACOH	1	1	0	0.9472136	0.2236068	0.0000000
10	CH2CO	2	1	0	1.6153551	0.5576775	0.1443376
11	CHO	1	1	0	0.9855986	0.2357023	0.0000000
12	FURFURAL	5	2	0	2.2774565	0.7237105	0.2276115
13	COOH	1	2	0	1.3554619	0.4277309	0.0912871
14	COOC	2	2	0	2.3164966	0.9082483	0.4915816
15	CH2O	1	1	0	1.1153551	0.2886751	0.0000000
16	CCl	1	0	0	2.1338934	1.1338934	0.0000000
17	CCl2	1	0	0	2.9748936	2.0874980	0.9091373
18	CCl3	1	0	0	3.9790305	1.9639610	2.2269225
19	CCl4	1	0	0	5.0355737	2.2677868	3.8571429
20	aCCl	1	0	0	1.6338934	0.5669467	0.0000000
21	CCN	2	0	1	1.4472136	0.4736068	0.1118034
22	ACNH2	1	0	1	1.0773503	0.2886751	0.0000000
23	CNO2	1	2	1	2.2247449	0.7415816	0.4013747
24	aCNO2	1	2	1	1.7247449	0.5374575	0.2347080
25	DOH	2	2	0	2.3086408	1.1324555	0.4472136
26	DEOH	4	3	0	4.1311026	2.2098058	1.0595860
27	PYRIDINE	5	0	1	1.9828944	0.6552630	0.2184210
28	TCE	2	0	0	4.4790305	2.0772222	1.6248376
29	MFA	2	1	1	2.4855986	1.0243774	0.4065263
30	DMF	3	1	1	3.4328122	1.3883283	1.0690206
31	TMS	4	2	0	2.2384584	0.6225212	0.2854595
32	DMSO	2	1	0	2.8164966	0.9831632	0.7415816

Table A.3. Stoichiometric and CIs values for the UNIFAC-CI (Dortmund) groups

Index	Group	nC	nO	nN	χ^0	χ^1	χ^2
1	CH2	1	0	0	0.7071068	0.0000000	0.0000000
2	C=C	2	0	0	1.0000000	0.2500000	0.0000000
3	ACH	1	0	0	0.5773503	0.6666667	0.1924501
4	ACCH2	2	0	0	1.2071068	0.3535534	0.0000000
5	OH	0	1	0	0.4472136	0.0000000	0.0000000

6	CH3OH	1	1	0	1.4472136	0.4472136	0.0000000
7	H2O	0	1	0	0.5000000	0.0000000	0.0000000
8	ACOH	1	1	0	0.7144748	0.2236068	0.0000000
9	CH2CO	2	1	0	1.6153551	0.5576775	0.1443376
10	CHO	1	1	0	0.9855986	0.2357023	0.0000000
11	CCOO	2	2	0	1.8164966	0.6582483	0.2874575
12	HCOO	1	2	0	1.3938469	0.4714045	0.0962250
13	CH2O	1	1	0	1.1153551	0.2886751	0.0000000
14	CNH2	1	0	1	1.5773503	0.5773503	0.0000000
15	CNH	1	0	1	1.5000000	0.5000000	0.0000000
16	(C)3N	1	0	1	1.4472136	0.4472136	0.0000000
17	ACNH2	1	0	1	0.8446115	0.1543033	0.0000000
18	PYRIDINE	5	0	1	1.6540117	0.4559422	0.1256795
19	CCN	2	0	1	1.4472136	0.4736068	0.1118034
20	COOH	1	2	0	1.3554619	0.4277309	0.0912871
21	CCl	1	0	0	2.1338934	1.1338934	0.0000000
22	CCl2	1	0	0	2.9748936	2.0874980	0.9091373
23	CCl3	1	0	0	3.9790305	1.9639610	2.2269225
24	CCl4	1	0	0	5.0355737	2.2677868	3.8571429
25	aCCl	1	0	0	1.6338934	0.5669467	0.0000000
26	CNO2	1	2	1	2.2247449	0.7415816	0.4013747
27	aCNO2	1	2	1	1.4920061	0.4424423	0.1571285
28	CS2	1	0	0	1.3164966	0.4082483	0.0833333
29	CH3SH	1	0	0	1.4472136	0.4472136	0.0000000
30	FURFURAL	5	2	0	2.2774565	0.7237105	0.2276115
31	DOH	2	2	0	2.3086408	1.1324555	0.4472136
32	NA	0	0	0	0.0000000	0.0000000	0.0000000
33	NA	0	0	0	0.0000000	0.0000000	0.0000000
34	C=-C	2	0	0	1.0000000	0.2500000	0.0000000
35	DMSO	2	1	0	2.8164966	0.9831632	0.7415816
36	ACRY	3	0	1	2.2316706	0.9205302	0.3332236
37	NA	0	0	0	0.0000000	0.0000000	0.0000000
38	ACF	1	0	0	0.6452257	0.2438724	0.0730850
39	DMF	3	1	1	3.4328122	1.3883283	1.0690206
40	CF2	1	0	0	1.2559289	0.3779645	0.0714286
41	COO	1	2	0	1.5245639	0.5468740	0.1290994
42	c-CH2	1	0	0	0.7071068	1.0000000	0.3535534
43	c-CH2O	2	1	0	1.1153551	0.5773503	0.2041241
44	HCOOH	1	2	0	1.4328122	0.4939012	0.1054093
45	CHCl3	1	0	0	3.4016803	1.9639610	2.2269225

APPENDIX B

ATOM INTERACTION PARAMETER TABLES FOR THE UNIFAC MODELS

Table B.1 AIPs corresponding to original UNIFAC-CI functional groups involving C, O, N, Cl and S atoms.

bC-C=	977.7980	dN-N=	-1919.6586	bhC-C=	-145.1011	dhN-N=	819.2628
cC-C=	-108.1096	eN-N=	-92.7458	chC-C=	-281.5677	ehN-N=	-784.2955
dC-C=	104.6162	bN-Cl=	0.0000	dhC-C=	321.4310	bhN-Cl=	0.0000
eC-C=	-109.4275	cN-Cl=	0.0000	ehC-C=	261.3034	chN-Cl=	0.0000
bC-O=	-1134.8172	dN-Cl=	0.0000	bhC-O=	399.9179	dhN-Cl=	0.0000
cC-O=	-17.4435	eN-Cl=	0.0000	chC-O=	-35.6315	ehN-Cl=	0.0000
dC-O=	45.2392	bN-S=	0.0000	dhC-O=	233.0069	bhN-S=	0.0000
eC-O=	61.8317	cN-S=	0.0000	ehC-O=	-144.3443	chN-S=	0.0000
bC-N=	-407.0000	dN-S=	0.0000	bhC-N=	-124.6281	dhN-S=	0.0000
cC-N=	-84.9200	eN-S=	0.0000	chC-N=	1302.1476	ehN-S=	0.0000
dC-N=	3592.6503	bCl-C=	0.0000	dhC-N=	72.2080	bhCl-C=	163.9330
eC-N=	-114.6263	cCl-C=	0.0000	ehC-N=	-783.4023	chCl-C=	-20.3295
bC-Cl=	-1292.0176	dCl-C=	0.0000	bhC-Cl=	0.0000	dhCl-C=	41.3890
cC-Cl=	1013.7207	eCl-C=	0.0000	chC-Cl=	0.0000	ehCl-C=	-219.2109
dC-Cl=	-33.3475	bCl-O=	0.0000	dhC-Cl=	0.0000	bhCl-O=	-1107.0263
eC-Cl=	438.0628	cCl-O=	0.0000	ehC-Cl=	0.0000	chCl-O=	-1686.5098
bC-S=	380.7717	dCl-O=	0.0000	bhC-S=	0.0000	dhCl-O=	2010.8859
cC-S=	845.1258	eCl-O=	0.0000	chC-S=	0.0000	ehCl-O=	15.6897
dC-S=	-731.9589	bCl-N=	0.0000	dhC-S=	0.0000	bhCl-N=	0.0000
eC-S=	198.8021	cCl-N=	0.0000	ehC-S=	0.0000	chCl-N=	0.0000
bO-C=	-318.5747	dCl-N=	0.0000	bhO-C=	63.3560	dhCl-N=	0.0000
cO-C=	-788.9818	eCl-N=	0.0000	chO-C=	-34.0054	ehCl-N=	0.0000
dO-C=	-781.6321	bCl-Cl=	0.0000	dhO-C=	195.5351	bhCl-Cl=	0.0000
eO-C=	363.0193	cCl-Cl=	0.0000	ehO-C=	-30.8774	chCl-Cl=	0.0000
bO-O=	637.0451	dCl-Cl=	0.0000	bhO-O=	-514.3574	dhCl-Cl=	0.0000
cO-O=	1126.7052	eCl-Cl=	0.0000	chO-O=	-78.1663	ehCl-Cl=	0.0000
dO-O=	477.6881	bCl-S=	60.2630	dhO-O=	98.2216	bhCl-S=	0.0000
eO-O=	-385.0802	cCl-S=	-793.0667	ehO-O=	-186.8762	chCl-S=	0.0000
bO-N=	-3570.3763	dCl-S=	-223.8041	bhO-N=	22.3762	dhCl-S=	0.0000
cO-N=	2991.3369	eCl-S=	65.8464	chO-N=	-108.8981	ehCl-S=	0.0000
dO-N=	-32.4800	bS-C=	0.0000	dhO-N=	66.4048	bhS-C=	11.4892
eO-N=	-1456.8287	cS-C=	0.0000	ehO-N=	-737.9977	chS-C=	-62.9930
bO-Cl=	1006.1964	dS-C=	0.0000	bhO-Cl=	0.0000	dhS-C=	-217.7392
cO-Cl=	-649.4863	eS-C=	0.0000	chO-Cl=	0.0000	ehS-C=	-160.5467
dO-Cl=	1272.5343	bS-O=	0.0000	dhO-Cl=	0.0000	bhS-O=	61.0982
eO-Cl=	198.5973	cS-O=	0.0000	ehO-Cl=	0.0000	chS-O=	47.4756
bO-S=	14.1513	dS-O=	0.0000	bhO-S=	0.0000	dhS-O=	128.1836
cO-S=	-515.7616	eS-O=	0.0000	chO-S=	0.0000	ehS-O=	302.4460
dO-S=	202.1089	bS-N=	0.0000	dhO-S=	0.0000	bhS-N=	0.0000
eO-S=	-15.4024	cS-N=	0.0000	ehO-S=	0.0000	chS-N=	0.0000
bN-C=	-312.9669	dS-N=	0.0000	bhN-C=	304.6581	dhS-N=	0.0000
cN-C=	97.8787	eS-N=	0.0000	chN-C=	-1518.3840	ehS-N=	0.0000
dN-C=	-1650.8984	bS-Cl=	0.0000	dhN-C=	2191.2183	bhS-Cl=	20.1149

eN-C=	394.7154	cS-Cl=	0.0000	ehN-C=	529.3639	chS-Cl=	66.7652
bN-O=	67.7593	dS-Cl=	0.0000	bhN-O=	-364.8061	dhS-Cl=	-168.6890
cN-O=	636.8919	eS-Cl=	0.0000	chN-O=	861.7200	ehS-Cl=	373.3153
dN-O=	193.4733	bS-S=	0.0000	dhN-O=	-2404.5812	bhS-S=	0.0000
eN-O=	1373.9865	cS-S=	0.0000	ehN-O=	-1506.4950	chS-S=	0.0000
bN-N=	-400.8128	dS-S=	0.0000	bhN-N=	110.6882	dhS-S=	0.0000
cN-N=	-173.3546	eS-S=	0.0000	chN-N=	1540.4155	ehS-S=	0.0000

Table B.2 AIPs corresponding to the original UNIFAC-CI molecular group CH_3OH related systems

bC-C=	417.1033	bO-C=	-66.4187	bhC-C=	14.0055	bhO-C=	86.4041
cC-C=	-22.3407	cO-C=	61.1223	chC-C=	-236.8398	chO-C=	-97.1444
dC-C=	-54.1940	dO-C=	-54.1938	dhC-C=	-250.5971	dhO-C=	-35.1994
eC-C=	-3.5115	eO-C=	-3.5115	ehC-C=	0.0000	ehO-C=	0.0000
bC-O=	396.9071	bO-O=	-74.8408	bhC-O=	-94.7931	bhO-O=	0.1451
cC-O=	-29.4132	cO-O=	93.4971	chC-O=	-236.8399	chO-O=	-97.1445
dC-O=	-94.6287	dO-O=	-94.6286	dhC-O=	-250.5971	dhO-O=	-35.1998
eC-O=	-14.4253	eO-O=	-14.4254	ehC-O=	0.0000	ehO-O=	0.0000

Table B.3 AIPs corresponding to the original UNIFAC-CI molecular group H_2O related systems

bC-C=	0.0000	bO-C=	-268.8366	bhC-C=	0.0000	bhO-C=	123.5371
cC-C=	0.0000	cO-C=	706.2878	chC-C=	0.0000	chO-C=	-1071.6053
dC-C=	0.0000	dO-C=	0.0000	dhC-C=	0.0000	dhO-C=	0.0000
eC-C=	0.0000	eO-C=	16.3036	ehC-C=	0.0000	ehO-C=	0.0000
bC-O=	-845.1770	bO-O=	963.8220	bhC-O=	-2205.5325	bhO-O=	909.4966
cC-O=	0.0000	cO-O=	71.5437	chC-O=	0.0000	chO-O=	5821.9204
dC-O=	0.0000	dO-O=	0.0000	dhC-O=	0.0000	dhO-O=	0.0000
eC-O=	0.0000	eO-O=	3.9220	ehC-O=	0.0000	ehO-O=	0.0000

Table B.4. AIPs corresponding to the original UNIFAC-CI molecular group PYRIDINE related systems

bC-C=	988.0659	dO-O=	0.0000	bhC-C=	-3.4234	dhO-O=	0.0000
cC-C=	-84.2139	eO-O=	0.0000	chC-C=	-242.2328	ehO-O=	0.0000
dC-C=	28.9377	bO-N=	0.0000	dhC-C=	402.4844	bhO-N=	0.0000
eC-C=	-44.8237	cO-N=	0.0000	ehC-C=	370.5852	chO-N=	0.0000
bC-O=	0.0000	dO-N=	0.0000	bhC-O=	359.9609	dhO-N=	0.0000
cC-O=	0.0000	eO-N=	0.0000	chC-O=	-50.9299	ehO-N=	0.0000
dC-O=	0.0000	bN-C=	-306.9166	dhC-O=	110.4087	bhN-C=	116.7745
eC-O=	0.0000	cN-C=	120.5933	ehC-O=	-51.7948	chN-C=	-1495.5216
bC-N=	-402.1131	dN-C=	-1632.5328	bhC-N=	-171.1524	dhN-C=	2222.4245
cC-N=	-77.3075	eN-C=	467.2085	chC-N=	1289.0657	ehN-C=	583.8862
dC-N=	3590.7041	bN-O=	0.0000	dhC-N=	-69.0159	bhN-O=	-362.3962
eC-N=	-98.8721	cN-O=	0.0000	ehC-N=	0.0000	chN-O=	888.7885
bO-C=	0.0000	dN-O=	0.0000	bhO-C=	0.0000	dhN-O=	-2398.9726
cO-C=	0.0000	eN-O=	0.0000	chO-C=	0.0000	ehN-O=	-1371.5786
dO-C=	0.0000	bN-N=	-401.3424	dhO-C=	0.0000	bhN-N=	114.7603
eO-C=	0.0000	cN-N=	-170.5513	ehO-C=	0.0000	chN-N=	1577.9300
bO-O=	0.0000	dN-N=	-1921.2043	bhO-O=	0.0000	dhN-N=	831.1490
cO-O=	0.0000	eN-N=	-79.9868	chO-O=	0.0000	ehN-N=	0.0000

Table B.5. AIPs corresponding to the original UNIFAC-CI molecular group CCl_4 related systems

bC-C=	958.7842	bN-C=	0.0000	bhC-C=	-103.7737	bhN-C=	0.0000
cC-C=	-124.8181	cN-C=	0.0000	chC-C=	-285.1413	chN-C=	0.0000
dC-C=	33.2628	dN-C=	0.0000	dhC-C=	318.5047	dhN-C=	0.0000
eC-C=	-126.6735	eN-C=	0.0000	ehC-C=	248.0737	ehN-C=	0.0000
bC-O=	0.0000	bN-O=	0.0000	bhC-O=	310.1377	bhN-O=	0.0000
cC-O=	0.0000	cN-O=	0.0000	chC-O=	0.0000	chN-O=	0.0000
dC-O=	0.0000	dN-O=	0.0000	dhC-O=	0.0000	dhN-O=	0.0000
eC-O=	0.0000	eN-O=	0.0000	ehC-O=	0.0000	ehN-O=	0.0000
bC-N=	0.0000	bN-N=	0.0000	bhC-N=	0.0000	bhN-N=	0.0000
cC-N=	0.0000	cN-N=	0.0000	chC-N=	0.0000	chN-N=	0.0000
dC-N=	0.0000	dN-N=	0.0000	dhC-N=	0.0000	dhN-N=	0.0000
eC-N=	0.0000	eN-N=	0.0000	ehC-N=	0.0000	ehN-N=	0.0000
bC-Cl=	-1310.9345	bN-Cl=	0.0000	bhC-Cl=	0.0000	bhN-Cl=	0.0000
cC-Cl=	951.1701	cN-Cl=	0.0000	chC-Cl=	0.0000	chN-Cl=	0.0000
dC-Cl=	-109.4461	dN-Cl=	0.0000	dhC-Cl=	0.0000	dhN-Cl=	0.0000
eC-Cl=	378.3183	eN-Cl=	0.0000	ehC-Cl=	0.0000	ehN-Cl=	0.0000
bO-C=	-321.6430	bCl-C=	0.0000	bhO-C=	0.0000	bhCl-C=	155.8860
cO-C=	-789.7447	cCl-C=	0.0000	chO-C=	0.0000	chCl-C=	-334.3415
dO-C=	-796.0563	dCl-C=	0.0000	dhO-C=	0.0000	dhCl-C=	355.6546
eO-C=	361.7188	eCl-C=	0.0000	ehO-C=	0.0000	ehCl-C=	-198.6543
bO-O=	0.0000	bCl-O=	0.0000	bhO-O=	0.0000	bhCl-O=	-1147.7694
cO-O=	0.0000	cCl-O=	0.0000	chO-O=	0.0000	chCl-O=	-1688.4220
dO-O=	0.0000	dCl-O=	0.0000	dhO-O=	0.0000	dhCl-O=	2057.6304
eO-O=	0.0000	eCl-O=	0.0000	ehO-O=	0.0000	ehCl-O=	-18.1662
bO-N=	0.0000	bCl-N=	0.0000	bhO-N=	0.0000	bhCl-N=	0.0000
cO-N=	0.0000	cCl-N=	0.0000	chO-N=	0.0000	chCl-N=	0.0000
dO-N=	0.0000	dCl-N=	0.0000	dhO-N=	0.0000	dhCl-N=	0.0000
eO-N=	0.0000	eCl-N=	0.0000	ehO-N=	0.0000	ehCl-N=	0.0000
bO-Cl=	1063.1271	bCl-Cl=	0.0000	bhO-Cl=	0.0000	bhCl-Cl=	0.0000
cO-Cl=	-550.6384	cCl-Cl=	0.0000	chO-Cl=	0.0000	chCl-Cl=	0.0000
dO-Cl=	1135.2316	dCl-Cl=	0.0000	dhO-Cl=	0.0000	dhCl-Cl=	0.0000
eO-Cl=	32.0017	eCl-Cl=	0.0000	ehO-Cl=	0.0000	ehCl-Cl=	0.0000

Table B.6. AIPs corresponding to the original UNIFAC-CI molecular group DOH related systems

bC-C=	1071.3331	bO-C=	-380.5842	bhC-C=	-167.4062	bhO-C=	41.0546
cC-C=	-77.9855	cO-C=	-817.1437	chC-C=	-290.2206	chO-C=	-42.6585
dC-C=	296.2914	dO-C=	-874.4455	dhC-C=	312.8363	dhO-C=	186.9403
eC-C=	-267.9259	eO-C=	435.6300	ehC-C=	299.1950	ehO-C=	7.0159
bC-O=	-1041.2821	bO-O=	575.0356	bhC-O=	401.5793	bhO-O=	-512.6959
cC-O=	12.6821	cO-O=	1098.5433	chC-O=	-35.8483	chO-O=	-78.3844
dC-O=	236.9142	dO-O=	384.8746	dhC-O=	232.8471	dhO-O=	98.0620
eC-O=	-96.6666	eO-O=	-312.4692	ehC-O=	-65.1710	ehO-O=	-107.7027

Table B.7. AIPs corresponding to the original UNIFAC-CI molecular group DMSO related systems

bC-C=	975.4933	dN-N=	0.0000	bhC-C=	-149.7776	dhN-N=	0.0000
cC-C=	-112.1958	eN-N=	0.0000	chC-C=	-277.5551	ehN-N=	0.0000
dC-C=	100.8302	bN-Cl=	0.0000	dhC-C=	300.7551	bhN-Cl=	0.0000
eC-C=	-114.4054	cN-Cl=	0.0000	ehC-C=	261.6412	chN-Cl=	0.0000
bC-O=	-1136.6443	dN-Cl=	0.0000	bhC-O=	399.6778	dhN-Cl=	0.0000
cC-O=	-20.1612	eN-Cl=	0.0000	chC-O=	-36.4885	ehN-Cl=	0.0000
dC-O=	44.9423	bN-S=	0.0000	dhC-O=	232.3875	bhN-S=	0.0000
eC-O=	58.6683	cN-S=	0.0000	ehC-O=	-145.4753	chN-S=	0.0000
bC-N=	0.0000	dN-S=	0.0000	bhC-N=	0.0000	dhN-S=	0.0000
cC-N=	0.0000	eN-S=	0.0000	chC-N=	0.0000	ehN-S=	0.0000
dC-N=	0.0000	bCl-C=	-0.3768	dhC-N=	0.0000	bhCl-C=	0.0000
eC-N=	0.0000	cCl-C=	-4.7145	ehC-N=	0.0000	chCl-C=	0.0000
bC-Cl=	0.0000	dCl-C=	-3.6199	bhC-Cl=	-5.0525	dhCl-C=	0.0000
cC-Cl=	0.0000	eCl-C=	-6.8855	chC-Cl=	3.0447	ehCl-C=	0.0000
dC-Cl=	0.0000	bCl-O=	0.7864	dhC-Cl=	-21.8811	bhCl-O=	0.0000
eC-Cl=	0.0000	cCl-O=	-1.3825	ehC-Cl=	-3.7082	chCl-O=	0.0000
bC-S=	378.9446	dCl-O=	-0.2879	bhC-S=	0.0000	dhCl-O=	0.0000
cC-S=	842.4080	eCl-O=	-2.4679	chC-S=	0.0000	ehCl-O=	0.0000
dC-S=	-732.2557	bCl-N=	0.0000	dhC-S=	0.0000	bhCl-N=	0.0000
eC-S=	195.6386	cCl-N=	0.0000	ehC-S=	0.0000	chCl-N=	0.0000
bO-C=	-319.0891	dCl-N=	0.0000	bhO-C=	65.0780	dhCl-N=	0.0000
cO-C=	-789.7007	eCl-N=	0.0000	chO-C=	-23.3729	ehCl-N=	0.0000
dO-C=	-782.9621	bCl-Cl=	0.0000	dhO-C=	193.8233	bhCl-Cl=	0.0000
eO-C=	362.1980	cCl-Cl=	0.0000	ehO-C=	-31.4791	chCl-Cl=	0.0000
bO-O=	636.5855	dCl-Cl=	0.0000	bhO-O=	-514.6612	dhCl-Cl=	0.0000
cO-O=	1126.1433	eCl-Cl=	0.0000	chO-O=	-79.0869	ehCl-Cl=	0.0000
dO-O=	476.5151	bCl-S=	61.0494	dhO-O=	97.4198	bhCl-S=	0.0000
eO-O=	-385.6932	cCl-S=	-794.4492	ehO-O=	-188.0709	chCl-S=	0.0000
bO-N=	0.0000	dCl-S=	-224.0921	bhO-N=	0.0000	dhCl-S=	0.0000
cO-N=	0.0000	eCl-S=	63.3786	chO-N=	0.0000	ehCl-S=	0.0000
dO-N=	0.0000	bS-C=	0.0000	dhO-N=	0.0000	bhS-C=	13.2112
eO-N=	0.0000	cS-C=	0.0000	ehO-N=	0.0000	chS-C=	-52.3606
bO-Cl=	0.0000	dS-C=	0.0000	bhO-Cl=	1.6310	dhS-C=	-219.4510
cO-Cl=	0.0000	eS-C=	0.0000	chO-Cl=	9.7281	ehS-C=	-161.1484
dO-Cl=	0.0000	bS-O=	0.0000	dhO-Cl=	-2.7347	bhS-O=	60.7944
eO-Cl=	144.7365	cS-O=	0.0000	ehO-Cl=	-4.5920	chS-O=	46.5550
bO-S=	13.6917	dS-O=	0.0000	bhO-S=	0.0000	dhS-O=	127.3817
cO-S=	-516.3235	eS-O=	0.0000	chO-S=	0.0000	ehS-O=	301.2513
dO-S=	200.9359	bS-N=	0.0000	dhO-S=	0.0000	bhS-N=	0.0000
eO-S=	-16.0155	cS-N=	0.0000	ehO-S=	0.0000	chS-N=	0.0000
bN-C=	0.0000	dS-N=	0.0000	bN-C=	0.0000	dhS-N=	0.0000
cN-C=	0.0000	eS-N=	0.0000	chN-C=	0.0000	ehS-N=	0.0000
dN-C=	0.0000	bS-Cl=	0.0000	dhN-C=	0.0000	bhS-Cl=	21.7459
eN-C=	0.0000	cS-Cl=	0.0000	ehN-C=	0.0000	chS-Cl=	76.4933
bN-O=	0.0000	dS-Cl=	0.0000	bhN-O=	0.0000	dhS-Cl=	-171.4237
cN-O=	0.0000	eS-Cl=	0.0000	chN-O=	0.0000	ehS-Cl=	368.7233
dN-O=	0.0000	bS-S=	0.0000	dhN-O=	0.0000	bhS-S=	0.0000
eN-O=	0.0000	cS-S=	0.0000	ehN-O=	0.0000	chS-S=	0.0000
bN-N=	0.0000	dS-S=	0.0000	bhN-N=	0.0000	dhS-S=	0.0000

cN-N=	0.0000	eS-S=	0.0000	chN-N=	0.0000	ehS-S=	0.0000

PARAMETER TABLES FOR LLE

Table B.8 AIPs corresponding to the original UNIFAC-CI (LLE) groups involving C, O, H and N atoms

bC-C=	995.0155	dO-O=	477.6881	bhC-C=	-146.7450	dhO-O=	98.2216
cC-C=	-120.8274	eO-O=	-246.6646	chC-C=	-288.4790	ehO-O=	-186.8762
dC-C=	99.5044	bO-N=	336.0519	dhC-C=	324.1394	bhO-N=	9.2323
eC-C=	-210.3474	cO-N=	1428.1509	ehC-C=	226.6315	chO-N=	-32.2790
bC-O=	-1112.2713	dO-N=	512.0932	bhC-O=	407.0801	dhO-N=	29.5052
cC-O=	-45.6407	eO-N=	-560.0939	chC-O=	-35.6315	ehO-N=	-231.9835
dC-O=	13.7526	bN-C=	0.0652	dhC-O=	233.0069	bhN-C=	200.6969
eC-O=	-34.4070	cN-C=	177.2827	ehC-O=	-144.3443	chN-C=	-186.0919
bC-N=	-1023.7077	dN-C=	194.3555	bhC-N=	-7.3726	dhN-C=	425.0065
cC-N=	-513.2711	eN-C=	-112.6086	chC-N=	269.2301	ehN-C=	31.8211
dC-N=	-85.2644	bN-O=	-5.5275	dhC-N=	-213.8597	bhN-O=	-278.8393
eC-N=	70.2626	cN-O=	-69.8627	ehC-N=	-246.6595	chN-O=	582.0711
bO-C=	-350.3757	dN-O=	-12.3291	bhO-C=	235.0258	dhN-O=	-96.8373
cO-C=	-803.2168	eN-O=	-134.5393	chO-C=	-43.7662	ehN-O=	-1193.1110
dO-C=	-781.6321	bN-N=	41.4208	dhO-C=	212.2180	bhN-N=	-66.6179
eO-C=	346.6379	cN-N=	301.1282	ehO-C=	-15.1433	chN-N=	209.9947
bO-O=	616.0487	dN-N=	318.5057	bhO-O=	-526.8468	dhN-N=	-377.7994
cO-O=	1162.8407	eN-N=	162.5345	chO-O=	-78.1663	ehN-N=	-246.6595

PARAMETER TABLE FOR MODIFIED UNIFAC-CI (DORTMUND)

Table B.9 AIPs corresponding to the UNIFAC-CI (Dortmund) groups involving C, O, H and N atoms

bC-C=	358.2658	d1O-O=	2.0393	bhC-C=	778.7618	dh1O-O=	1.7971
cC-C=	-52.3157	e1O-O=	0.1460	chC-C=	-150.3735	eh1O-O=	1.1447
dC-C=	-128.0109	b1O-N=	0.0538	dhC-C=	-97.4092	bh1O-N=	0.0000
eC-C=	-1.6762	c1O-N=	-0.0012	ehC-C=	-56.5978	ch1O-N=	0.0000
bC-O=	-5133.7362	d1O-N=	0.0000	bhC-O=	-1766.2730	dh1O-N=	0.0000
cC-O=	-1810.9141	e1O-N=	-0.2341	chC-O=	3978.4731	eh1O-N=	0.0000
dC-O=	-89.2293	d1N-C=	-0.0233	dhC-O=	2446.0265	dh1N-C=	-0.0516
eC-O=	-908.6551	e1N-C=	-0.1246	ehC-O=	420.8144	eh1N-C=	-0.7278
bC-N=	-552.8644	b1N-C=	-0.2471	bhC-N=	-7.4304	bh1N-C=	0.7164
cC-N=	-811.7053	c1N-C=	-0.7549	chC-N=	269.2800	ch1N-C=	0.3585
dC-N=	-475.8557	d1N-O=	0.0000	dhC-N=	-214.0393	dh1N-O=	0.8091
eC-N=	122.2891	e1N-O=	0.0000	ehC-N=	-246.6595	eh1N-O=	0.0000
bO-C=	1423.2157	b1N-O=	0.0000	bhO-C=	714.9907	bh1N-O=	0.0000
cO-C=	91.8698	c1N-O=	0.0000	chO-C=	-19.6798	ch1N-O=	0.0000
dO-C=	1192.5284	d1N-N=	-0.0128	dhO-C=	-248.5684	dh1N-N=	0.1484

eO-C=	136.6679	e1N-N=	-0.1976	ehO-C=	-256.6972	eh1N-N=	-1.9677
bO-O=	-57.7635	d1N-N=	0.0676	bhO-O=	1051.6161	dh1N-N=	0.3081
cO-O=	-1494.8288	e1N-N=	-0.4298	chO-O=	4674.4442	eh1N-N=	0.0000
dO-O=	-222.8789	b2C-C=	0.0000	dhO-O=	-178.9704	bh2C-C=	0.0000
eO-O=	-323.6611	c2C-C=	0.0000	ehO-O=	-251.9237	ch2C-C=	0.0000
bO-N=	365.3837	d2C-C=	0.0000	bhO-N=	9.2323	dh2C-C=	0.0000
cO-N=	1434.2510	e2C-C=	0.0000	chO-N=	-32.2790	eh2C-C=	0.0000
dO-N=	512.0932	b2C-O=	0.0000	dhO-N=	29.5052	bh2C-O=	0.0000
eO-N=	-576.6557	c2C-O=	0.0000	ehO-N=	-231.9835	ch2C-O=	0.0000
bN-C=	0.0468	d2C-O=	0.0000	bhN-C=	-121.2320	dh2C-O=	0.0000
cN-C=	177.1865	e2C-O=	0.0000	chN-C=	-178.3640	eh2C-O=	0.0000
dN-C=	194.2954	b2C-N=	0.0000	dhN-C=	422.0607	bh2C-N=	0.0000
eN-C=	-113.0743	c2C-N=	0.0000	ehN-C=	69.6468	ch2C-N=	0.0000
bN-O=	-5.5275	d2C-N=	0.0000	bhN-O=	-353.3968	dh2C-N=	0.0000
cN-O=	-69.8627	e2C-N=	0.0000	chN-O=	582.0711	eh2C-N=	0.0000
dN-O=	-12.3291	b2O-C=	0.0000	dhN-O=	-96.8373	bh2O-C=	0.0000
eN-O=	-134.5393	c2O-C=	0.0000	ehN-O=	-1193.1110	ch2O-C=	0.0000
bN-N=	41.4201	d2O-C=	0.0000	bhN-N=	-66.6199	dh2O-C=	0.0000
cN-N=	301.0891	e2O-C=	0.0000	chN-N=	210.1005	eh2O-C=	0.0000
dN-N=	318.5027	b2O-O=	0.0000	dhN-N=	-377.8084	bh2O-O=	0.0000
eN-N=	162.3107	c2O-O=	0.0000	ehN-N=	-246.6595	ch2O-O=	0.0000
b1C-C=	-2.9444	d2O-O=	0.0000	bh1C-C=	-3.4265	dh2O-O=	0.0000
c1C-C=	0.2408	e2O-O=	0.0000	ch1C-C=	0.4320	eh2O-O=	0.0000
d1C-C=	0.3989	b2O-N=	0.0000	dh1C-C=	0.2966	bh2O-N=	0.0000
e1C-C=	0.0084	c2O-N=	0.0000	eh1C-C=	0.1531	ch2O-N=	0.0000
b1C-O=	9.5931	d2O-N=	0.0000	bh1C-O=	4.8099	dh2O-N=	0.0000
c1C-O=	12.6613	e2O-N=	0.0000	ch1C-O=	-12.8848	eh2O-N=	0.0000
d1C-O=	0.1564	b2N-C=	0.0000	dh1C-O=	-3.9140	bh2N-C=	0.0000
e1C-O=	1.1511	c2N-C=	0.0000	eh1C-O=	-0.1915	ch2N-C=	0.0000
b1C-N=	13.1284	d2N-C=	0.0000	bh1C-N=	0.9680	dh2N-C=	0.0000
c1C-N=	13.8454	e2N-C=	0.0000	ch1C-N=	-1.0701	eh2N-C=	0.0000
d1C-N=	13.5740	b2N-O=	0.0000	dh1C-N=	3.6418	bh2N-O=	0.0000
e1C-N=	-0.4885	c2N-O=	0.0000	eh1C-N=	0.0000	ch2N-O=	0.0000
b1O-C=	-5.9784	d2N-O=	0.0000	bh1O-C=	-2.2397	dh2N-O=	0.0000
c1O-C=	-1.7431	e2N-O=	0.0000	ch1O-C=	0.0754	eh2N-O=	0.0000
d1O-C=	4.1701	b2N-N=	0.0000	dh1O-C=	0.7525	bh2N-N=	0.0000
e1O-C=	-0.1751	c2N-N=	0.0000	eh1O-C=	0.7973	ch2N-N=	0.0000
b1O-O=	0.3022	d2N-N=	0.0000	bh1O-O=	-3.4953	dh2N-N=	0.0000
c1O-O=	0.4037	e2N-N=	0.0000	ch1O-O=	-17.3469	eh2N-N=	0.0000

APPENDIX C

Parameter Estimation Technique

The non-linearity in the UNIFAC equations has been studied through different approaches (see, for example, the work of McDonald et al.[93], Kratch et al.[85]) This work was devoted to estimation of the needed thermodynamic model parameters. Typically, thermodynamic model parameters are estimated via a sum of weighted squares (Kemeny et al.[94]); several authors have used this approach for parameter estimation when developing thermodynamic models[95]. It has been demonstrated that the sum of weighted squares is an efficient technique for estimation of thermodynamic model parameters due to the fact that the of the optimization scenario inherent to the development of the UNIFAC-CI models is very complex; it has been decided to use the sum of weighted squares to estimate parameters in this work.

When regressing a parameter in a non-linear model to experimental data, the following is required:

1) A numerical rule for successively updating iterates;
2) A method for deciding when to stop the process; and
3) Starting values to get the iterative process under way.

The idea behind the sum of least squares (sometimes termed 'sum of squares', S) principle, when matching experimental data, is the minimization of the sum of squared deviations between the experimental observations and the calculated values. The basic model equation is:

$$Y_i = f\left(x_i, \theta\right) + e_i, \quad ei \sim iid\left(0, \sigma^2\right),\ i = 1, \cdots, n \tag{C.1}$$

where θ is the ($p \times 1$) vector of parameters to be estimated, and $f(x_i, \theta)$ is the mean of Y_i. e_i are the random errors or disturbances; and it is supposed to be independent, identically distributed (*iid*), mean 0 and variance σ^2.
The residual sum of squares to be minimized can now be written as:

$$\begin{aligned} S(\theta) &= \sum_{i=1}^{n} \left(yi - f\left(x_i, \theta\right) \right)^2 \\ &= \left(y - f\left(x, \theta\right) \right)' \left(y - \left(x, \theta\right) \right), \end{aligned} \tag{C.2}$$

with

$$f(x, \theta) = \begin{bmatrix} f\left(x_1, \theta\right) \\ f\left(x_2, \theta\right) \\ f\left(x_3, \theta\right) \end{bmatrix} \tag{C.3}$$

A Newton´s method is chosen to find the minimum, using a Levenberg-Marquardt type algorithm to force the Newton method in the direction of steepest descent. The algorithm is implemented[1] in the Harwell FORTRAN subroutine VA07AD[95], which has been used as minimization algorithm. When the process has successfully converged, the converged iterate is called the non-linear least squares estimates $\hat{\theta}$ of θ.

The vector of starting values, θ^0, is very important, as it was stated in Chapter 4. The closer the starting values are to the least squares estimate that minimizes the objective function, the faster the convergence and the more reliable the iterative algorithm.

APPENDIX D

Correlation Statistics for the UNIFAC-CI LLE Model

The present Appendix shows the correlation results for the UNIFAC-CI LLE model in comparison with the reference model (UNIFAC-LLE). The deviation results are sorted in ascending way with respect to the CI model. Measures of deviation are defined as follow:

$$Deviation\ (DVSO) = \frac{1}{ND}\sum_{i=1}^{ND}\left|\frac{\left(x_i^{\exp} - x_i^{calc}\right)}{x_i^{\exp}}\right| \times 100 \qquad (C.3)$$

where

ND : Number of data points in the data set
$x_i^{\exp}$: Measured mol fraction composition
x_i^{calc} : Calculated mol fraction composition

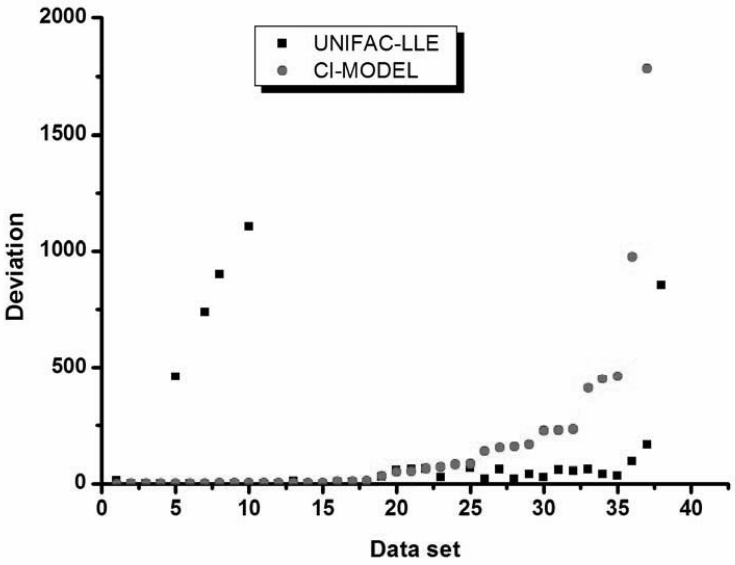


Table D.1. Correlation statistics. Deviations of the individual data sets for UNIFAC-CI compared against original UNIFAC, where DS is number of data set.

DS	SYSTEM	UNIFAC	CI-MODEL
1	JA_707_2-METHYL-1-PROPANOL_WATER.LLE	13.896	0.406
2	JA_534_CYCLOPENTANE_WATER.LLE	0.03	1.722
3	JA_542_CYCLOHEXANE_WATER.LLE	0.05	2.188
4	JA_586_CYCLOHEXANE_WATER.LLE	0.043	2.218
5	JA_331_PENTANE_WATER.LLE	457.35	2.34
6	JA_1_PENTANE_WATER.LLE	0.025	2.4825
7	JA_332_HEXANE_WATER.LLE	737.48	2.515
8	JA_333_HEPTANE_WATER.LLE	899.27	2.706
9	JA_6_HEPTANE_WATER.LLE	0.005	2.885
10	JA_334_OCTANE_WATER.LLE	1104.56	2.9
11	JA_522_HEPTANE_WATER.LLE	0.045	3.12
12	JA_263_HEPTANE_WATER.LLE	0.03	3.14
13	JA_661_1-PENTANOL_WATER.LLE	12.246	3.242
14	JA_264_NONANE_WATER.LLE	0.04	3.6
15	321_NITROBENZENE_WATER.LLE	1.8583	5.25
16	112_NITROBENZENE_WATER.LLE	8.76	10.42
17	233_PROPIONITRILE_WATER.LLE	6.9357	10.75
18	JA_490_1-DODECANOL_WATER.LLE	5.57	12.64
19	644_PROPIONITRILE_WATER.LLE	30.82	34.435
20	1035_NITROMETHANE_WATER.LLE	57.575	51.295
21	JA_680_CYCLOHEXANOL_WATER.LLE	62.58	54.046
22	313_ADIPONITRILE_WATER.LLE	63.186	64.499
23	JA_781_2-PENTANOL_WATER.LLE	26.45	71.395
24	JA_238_HEXANE_WATER.LLE	2654.58	83.26
25	JA_625_2-BUTANOL_WATER.LLE	68.25	85.51
26	JA_158_2-HEXANOL_WATER.LLE	21.92	139.29
27	JA_206_1-BUTANOL_WATER.LLE	59.91	155.13
28	JA_357_1-PENTANOL_WATER.LLE	21.535	157.725
29	JA_681_1-BUTANOL_WATER.LLE	40.907	168.65
30	JA_510_1-PENTANOL_WATER.LLE	29.31	226.4522
31	JA_753_2-METHYL-1-PROPANOL_WATER.LLE	58.85	229.098
32	JA_427_1-BUTANOL_WATER.LLE	54.54	233.205
33	JA_468_1-BUTANOL_WATER.LLE	61.62	407.77
34	JA_429_2-METHYL-1-PROPANOL_WATER.LLE	41.42	447.7056
35	JA_359_1-BUTANOL_WATER.LLE	35.58	459.17
36	JA_360_1-HEXANOL_WATER.LLE	95.53	974.14
37	JA_361_1-HEPTANOL_WATER.LLE	166.55	1782.88

The discrepancy in the deviation values (even for the same system) are due the fact that deviations were sometimes calculated either for very dilute regions (which increases the numerical values of the deviations) and along all the concentration range. Then, the calculations of the deviations are sometimes of very different orders of magnitude when deviations for diluted and non-diluted systems are compared.

This PhD-project was carried out at CAPEC, the Computer Aided Product-Process Engineering Center.

CAPEC is committed to research, to work in close collaboration with industry and to participate in educational activities. The research objectives of CAPEC are to develop computer-aided systems for product/process simulation, design, analysis and control/operation for chemical, petrochemical, pharmaceutical and biochemical industries. The dissemination of the research results of CAPEC is carried out in terms of computational tools, technology and application. Under computational tools, CAPEC is involved with mathematical models, numerical solvers, process/operation mathematical models, numerical solvers, process simulators, process/product synthesis/design toolbox, control toolbox, databases and many more. Under technology, CAPEC is involved with development of methodologies for synthesis/design of processes and products, analysis, control and operation of processes, strategies for modelling and simulation, solvent and chemical selection and design, pollution prevention and many more. Under application, CAPEC is actively involved with developing industrial case studies, tutorial case studies for education and training, technology transfer studies together with industrial companies, consulting and many more.

Further information about CAPEC can be found at www.capec.kt.dtu.dk.

Department of Chemical
and Biochemical Engineering
DTU Building 229
Søltofts Plads
DK-2800 Kgs. Lyngby
www.kt.dtu.dk